NOTES ON ELLIPTIC CURVES

MATH 3320, LARRY ROLEN

1. Motivating examples

In order to understand the cryptographic applications, we first need to learn a little bit about elliptic curves in general. These are classical objects in number theory, which are closely tied to some of the oldest problems in mathematics. We first mention two of the most famous of these applications.

1.1. **Congruent Numbers.** We say that a natural number n is **congruent** if it is the area of a rational right triangle (that is, with rational side lengths). The question of which numbers are congruent goes back a few thousand years, to ancient Greek and Arab mathematicians, and much later to Fibonacci and later still Fermat. For example, 6 is the area of a $3-4-5$ triangle, so 6 is congruent. Pythagorean triples are easy to characterize (this was already done in the Elements), but if we allow *rational* side lengths, it is not a finite search to check through all possible denominators to determine whether a number is congruent or not. Fermat knew how to give proofs by contradiction that numbers like $1, 2, 3$ are not congruent. However, until recently, there was no easy way to quickly check whether a given n is congruent or not.

To describe this, we begin with a strange-looking, but yet elementary, lemma. First, note that the geometric condition for n to be congruent is that there are $X, Y, Z \in \mathbb{Q}$ such that

$$X^2 + Y^2 = Z^2, \qquad \frac{XY}{2} = n.$$

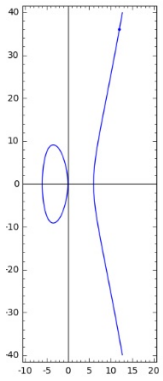
The lemma is purely a set of algebraic identities, so we omit the proof (as, for the sake of time, we have also omitted the intuitive explanation of where this "strange" lemma comes from).

Lemma 1.1. *There is a one-to-one correspondence*

$$\left\{(X,Y,Z) \in \mathbb{Q}^3 : X^2 + Y^2 = Z^2, XY/2 = n\right\} \leftrightarrow \{(x,y) \in \mathbb{Q}^2 : y^2 = x^3 - n^2 x,\ y \neq 0\}$$

given explicitly by $(X,Y,Z) \mapsto (nY/(Z-X), 2n^2/(Z-X))$ *in one direction and* $(x,y) \mapsto ((x^2 - n^2)/y, 2nx/y, (x^2+n^2)/y)$ *in the other.*

Note that the equation $y^2 = x^3 - n^2 x$ has 3 "obvious" solutions in the rationals: when $y = 0$, there are the solutions with $x = 0, \pm n$. These, however, are exactly the ones left out of our correspondence! Note also in the correspondence that in one direction we divide by factors of $Z - X$; these are never 0 as the hypotenuse of a right triangle must have length strictly larger than the lengths of the legs. Here is a numerical example. Our first example of a congruent number is $n = 6$, which is the area of the $3-4-5$ Pythagorean triangle. If we plug this into the explicit formulas in the lemma, we have $(X,Y,Z) = (3,4,5)$, and so $(x,y) = (6 \cdot 4/2, 2 \cdot 62/2) = (12, 36)$. Thus, the curve $y^2 = x^3 - 36x$ has the rational point $(12, 36)$ on it:



The curve $y^2 = x^3 - n^2x$ is our first example of an *elliptic curve*. Below, we will discuss the general story of elliptic curves, and we shall find that they have an extra special "structure" on them which has a lot to tell us about Diophantine problems like the congruent number problem.

Even when a number is congruent, it is often impossible to find an actual triangle representing it as such. For example 157 is congruent, and Don Zagier famously computed that the "simplest" triangle showing this has hypotenuse: $\frac{224403517704336969924557513090674863160948472041}{8912332268928859588025535178967163570016480830}$.

Today, we have a fast answer for the congruent number problem, but it is only conjecturally true, assuming a famous $1 Million problem.

1.2. **Fermat's Last Theorem.** Another important Diophantine equation, which we've already studied is that which shows up in Fermat's Last Theorem. Recall that this claims that there are no positive integral solutions to the equation

$$x^n + y^n = z^n, \ n \geq 3.$$

Simple proofs by contradiction (pioneered by Fermat) can be used for some small cases, like $n = 4$. we saw a hint of why this problem is extremely difficult (because, unfortunately, the "Fundamental Theorem of Arithmetic" fails for more general number sets than the integers). We now know, thanks to a deep and difficult proof of Wiles, obtained more than 350 years after Fermat first formulated the problem. The key idea behind Wiles' proof relies in an essential way on elliptic curves. Explicitly, an idea proposed by Frey works as follows. Suppose that there were a counter-example to Fermat's Last Theorem, given by a triple (a, b, c). Frey noticed that in such a situation, say when $x^p + y^p = z^p$ (its easy to check that its enough to prove FLT for prime exponents and for $n = 4$ which we know, so we assume here that p is a prime), then we can define a plane curve, called a *Frey curve*, by the equation

$$y^2 = x(x - a^p)(x - b^p).$$

This will turn out, after we've discussed the actual definition, to be an elliptic curve. What was noticed, and later proven by Wiles with help from a result of Serre and Ribet, is that this elliptic curve would have very strange properties. In fact, these properties are so strange, that you suspect such a curve shouldn't exist. Moreover, elliptic curves were also conjectured to be closely related to another mathematical object, called *modular forms*. Serre and Ribet showed that if an elliptic curve is closely related to a modular form, which was shown by Wiles to be true in the cases needed here, then these strange properties cannot occur. Finally, many other Diophantine equations can be studied via similar elliptic curves and Wiles-style methods. For instance, a relatively recent

Annals paper of Bugeaud, Mignotte, and Siksek applied similar methods to show that the only Fibonacci numbers which are perfect powers are $0, 1, 8, 144$, which is a shockingly simple, if not deceptively so, result.

1.3. **Cryptography.** We have discussed some important cyrptoschemes, such as RSA. Increasingly important cyrptoschemes, which have many advantages (small key sizes, harder to attack than RSA using methods of attack like we saw) and are becoming national standards in many arenas, are based on elliptic curves. Just like RSA, there is a mathematical operation which is easy to compute in one direction and hard to invert, which underpins the cyrptography. With RSA, this was multiplication of primes vs. factorization into primes. We have seen another such example recently; namely, primitive roots modulo n. Its easy to take a power of an integer mod something, but hard to compute *discrete logarithm*, the power of the given primitive root corresponding to any given invertible congruence class. Like $\mathbb{Z}_n^*$, elliptic curves have a group structure (this is the same structure hinted at above), and this can be used for cyrptography via the same kind of problem of computing discrete logarithms, just with a different group. For example, if you use Firefox, by playing around in the advanced settings, you can view the information on the elliptic curves it is using to encrypt your data. The chip on your credit card is also able to perform such elliptic curve computations to secure your privacy.

2. Elliptic Curves

2.1. **Definitions and Examples.** Ellpitic curves will be, first of all, curves, and specifically curves satisfying an equation of the form:

$$y^2 = x^3 + Ax + B$$

for some numbers A, B. We will have to wait a moment to see "why" equations of precisely this shape are so special. We only want to study curves which don't have any "bad points", or singularities. Just as you saw in Calculus class, some curves have "sharp points", or cusps. For example, the function $y = |x|$ has such a point at the origin, and we don't like this point, as the function is not differentiable there (or, if a particle were to travel along this curve at a constant speed with respect to arc length, it would experience an "infinite" acceleration at that point due to the sudden change of direction, and this never happens in "real life").

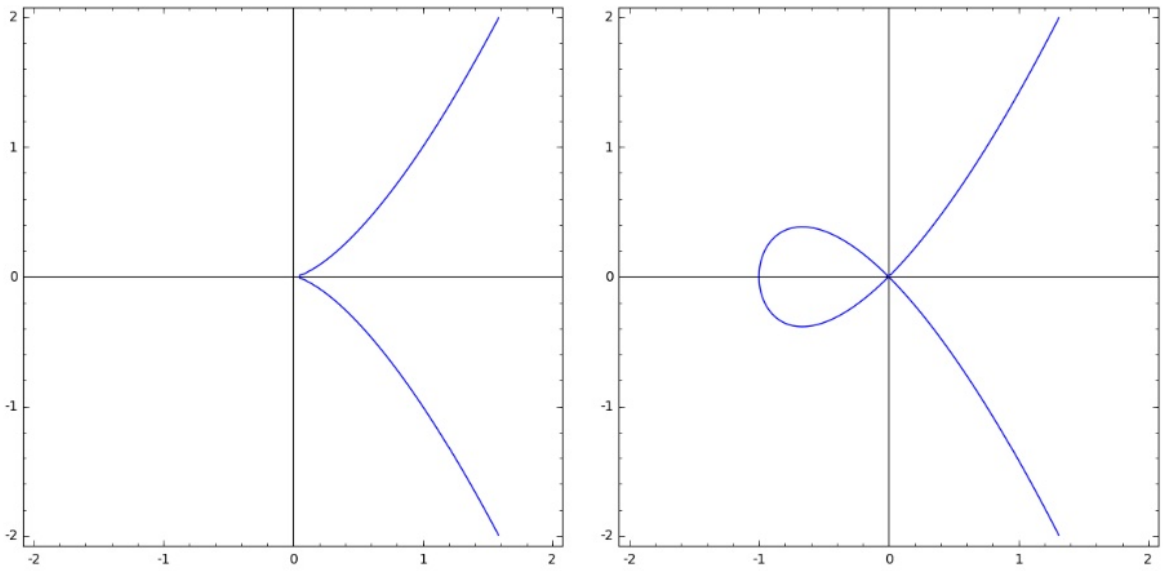
In this instance (and, crucially, to unravel the special structure of elliptic curves which makes them so useful), we need to introduce projective coordinates. You have possibly heard of the idea of perspective in art or vision, where two parallel lines "look like" they converge very very far away to the same place. For instance, the two rails on a set of train tracks appear to meet off in the distance. There is a related idea in mathematics, where two parallel lines do intersect, just very far away, at "infinity." Roughly speaking, why this is useful here is that if you took two different random lines, which are defined by polynomial equations of degree 1, they almost always intersect exactly once, except for the degenerate case of parallel lines (which has a 0% chance of happening). However, we'd like to count those as intersecting at one point as well, and if we use projective coordinates, curves defined by a polynomial of degree m will always intersect curves defined by polynomials of degree n exactly the nice and predictable number mn times.

We can formally describe this as follows. The (complex) projective plane is the set of equivalence relations of triples $(x, y, z) \in \mathbb{C}^3 \setminus \{(0, 0, 0)\}$ under the equivalence relation $(x, y, z) \sim (\lambda x, \lambda y, \lambda z)$ for each $\lambda \in \mathbb{C}$. If we take a representative of a point in the projective plane, we write it as $(x : y : z)$, and we call this representation a set of homogenous coordinates. This is really an

extension of an ordinary plane, of ordinary tuples of complex numbers (x, y), as if $z \neq 0$ for some point in the projective plane, then we can WLOG rescale so that z = 1 and obtain a representative of the form $(x : y : 1)$, which directly corresponds to the ordered pair (x, y). What we gain by taking the projective plane, is the additional "points at infinity," where $z = 0$. This is called the *line at infinity.*

Note that a curve defined by a polynomial in the projective plane must be homogenous in order to be compatible with the rescaling operation. If we have any ordinary plane curve, we can just "homogenize" the equation by inserting a power of z in each monomial factor until the maximal degree of the original polynomial matches the total degree of each monomial. For instance, for the lines $y = x + 1$ and $y = x - 2$, which are parallel, we rewrite these as the "zero-sets" of $y - x - 1$ and $y - x + 2$, and homogenize to obtain $y - x - z$ and $y - x + 2z$. Incidentally, we can check that these parallel lines now do intersect, as if we subtract the first equation from the second, we obtain $3z = 0$ and so $z = 0$ (so they do only intersect "at infinity"), and plugging back into the original equations we find that $x = y$. Since $(0, 0, 0)$ isn't an allowable point, we have $x, y \neq 0$, and we can thus rescale to find that the lines intersect at infinity at exactly one point, namely $(1 : 1 : 0)$.

What about for a cubic equation $y^2 = x^3 + Ax + B$? Projectivizing this gives the equation $y^2z - x^3 - Axz^2 + Bz^2 = 0$. What are the points at infinity? Well, if $z = 0$, then we obtain $-x^3 = 0$, and so $x = 0$. Now y isn't also allowed to be 0, so this cubic curve always has precisely one point at infinity, namely the point $(0 : 1 : 0)$. There are two types of bad behavior that can happen for such cubic curves, illustrated in the following picture:



These illustrate (from left to right) the curves $y^2 = x^3$ and $y^2 = x^3 + x^2$. The first has a *cusp*, and the second also has a problem at the origin, where the curve intersects itself (called a *node*). What is going on here is that a curve defined by the polynomial equation $f(x_1, x_2, \ldots, x_n) = 0$ has a singular point at P on the curve if the partial derivatives at this point **all** vanish at that point. For example, in the case of $y^2 = x^3 + x^2$, we write this as the zero set of the projectivized equation $y^2z - x^3 - x^2z$ which has partial derivatives with respect to x, y, z : $-3x^2 - 2xz, 2yz, y^2 - x^2$, which

all vanish at the point $(0:0:1)$ (corresponding to the origin in the picture above). If a plane curve has no singularities, it is called *smooth*.

Finally, we can say that the equation $E\colon y^2 = x^3 + Ax + B$ (we can have quadratic and other terms as in the example above, but they can always be converted to an equation of this form by a change of variables) is an *elliptic curve* if it is smooth. For which A, B does this hold? Well, we can write the projectivized version as $f(x, y, z) = y^2z - x^3 - Axz^2 - Bz^3$, which has partial derivatives $-3x^2 - Az^2, 2yz$, and $y^2 - 2Axz - 3Bz^2$. The curve is always smooth at the point at infinity, as if $z = 0$ and these three quantities are all 0, then the first equation implies that $x = 0$, and then the third equation implies that $y = 0$, which isn't allowed. Thus, we can assume that $z = 1$, we solve for the partial derivatives of $f(x, y, 1) = y^2 - x^3 - Ax - B$ being 0:

$$f_x = -3x^2 - A = f_y = 2y = 0.$$

The second implies that $y = 0$, and so the singular points on the curve are where $A = -3x^2$ and $x^3 + Ax + B = 0$ (so that its actually on the curve). Plugging the first into the second gives $x^3 - 3x^3 + B = -2x^3 + B = 0$, or $x^3 = B/2$. Thus, $(B/2)^2 = -(A/3)^3 = x^6$, and so $4A^3 + 27B^2 = 0$. This quantity, $\Delta(E) = \Delta := 4A^3 + 27B^2$, is called the *discriminant* of the elliptic curve. The answer to our question above, is then that a plane cubic $y^2 = x^3 + Ax + B$ is smooth, and hence an elliptic curve, exactly when $\Delta \neq 0$.

This begs the question, though, of what makes elliptic curves special, and why we would make a definition as above. Soon, we will see that elliptic curves have a very special, and rare, extra structure which gives us a lot of extra information. Curves described by equations of lower degree, like plane conics, can in many cases have all of their solutions completely parameterized (like for Pythagorean triples, which are rational points on a circle), and that looking for congruence and sign obstructions gives us all instances where there are no rational points. Curves of higher degree certainly show up in interesting Diophantine questions (like Fermat's Last Theorem!), but many things are much harder to determine for high degree curves (even questions like is there a rational solution are not so easy, let alone quantifying how many there are in a given range, even though many problems in number theory can be phrased in terms of such questions). Thus, elliptic curves are an intermediate "Goldilocks" case, between examples like plane conics which are known very explicitly, and higher degree curves where many things are still not known. That is, it is not as though all natural Diophantine equations have to do with elliptic curves, but rather that when we are lucky and they do, we can say a lot more about such problems.

3. The group structure of elliptic curves

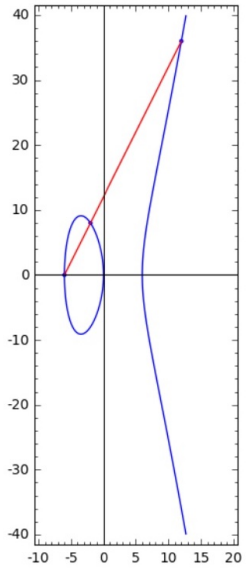
We saw that the set of rational points on the elliptic curve $y^2 = x^3 - n^2x$ determines whether n is congruent or not. If $E\colon y^2 = x^3 + Ax + B$ is an elliptic curve, and F is a field, we denote by $E(F)$ the set of points on E with coordinates in F; that is,

$$E(F) := \{(x, y) \in F^2 : y^2 = x^3 + Ax + B\}.$$

For example, the set of rational points (x, y) on E is denoted by $E(\mathbb{Q})$. What makes elliptic curves so special is that this set is in fact a **group** under a special operation. We will first see how this works geometrically, and then see that we can write it *algebraically*, which is key for the cryptography applications. First, we need a key result.

Theorem (Bezout)**.** *Two projective plane curves with no common component (think: e.g. they aren't the same curve, so they don't intersect in infinitely many points, and specifically, if you compute the gcd of their two polynomials, it must be constant), of degrees m, n (=degree of the defining polynomial) intersect exactly mn times (with multiplicity).*

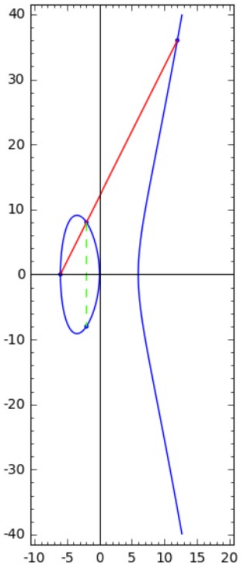
As for what multiplicity means, as an example, if a line intersects a curve, it usually has multiplicity 1 at a point of intersection, but if it is tangent to the curve, then it has multiplicity greater than 1. Now group laws on sets take two elements and combine them to give a third element. Elliptic curves are plane curves of degree 3. Thus, if we have two points on an elliptic curve, and we'd like to "add" them, then there is a unique line between these two points. But a line has degree 1, and by Bezout's Theorem, it intersects an elliptic curve in 3 points, two of which are already accounted for. Here is a picture of this procedure in action. This is for the curve $y^2 = x^3 - 36x$ we saw above in relation to the congruent number problem, together with the two points $(-6, 0)$ and $(12, 36)$ on the curve we saw above, and the line $y = 2(x + 6)$, which intersects the curve at $(-2, 8)$:



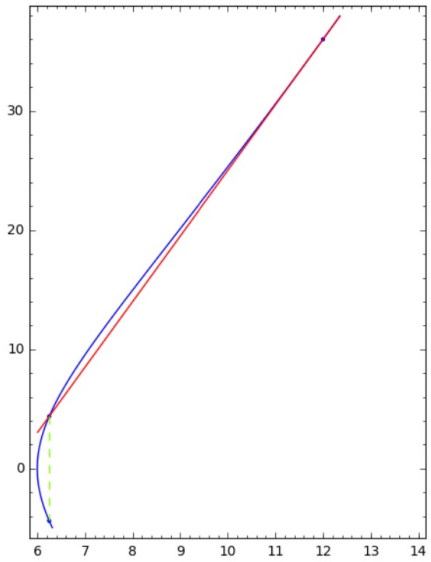
Thus, we have input two points on the curve, and obtained a third point on the curve. However, the "sum" of our two points won't be this point, but we are close. To describe the group law, first we must fix a point, any point, on the curve. It is convenient and standard to pick the point at infinity from above as the identity element of our group (playing the role that 0 plays in the group of real numbers under addition), which we then call "0." If we pick another point P on the curve, what should $-P$ be? Well, it should be a point such that when you draw a line through P and $-P$ the third intersection point guaranteed by Bezout's theorem is at infinity. From the pictures of elliptic curves above (they are symmetric around the x-axis as the only dependence on y is invariant under $y \mapsto -y$, we can see that the reflection of P across the x-axis must be $-P$.

Thus, if you add P to 0, the line through 0,P,$-P$ is the vertical line through P. If you consider the points you want to add as 0 and P, then the third point of intersection on this line is $-P$, the reflection of P. But we require that $P + 0 = P$, so we have to draw the line connecting P and 0, then intersect with a third point, and then reflect across the x-axis to get back to P. In general, to

add any two points P and Q, we draw the line and intersect with the curve as above, then reflect, so that in the case of $y^2 = x^3 - 36x$, $(-6, 0) + (12, 36) = (-2, -8)$, as pictured:



We have ignored one small issue in this procedure. What if we want to add a point P to itself? Well, then we draw a tangent line to the curve at P (which intersects the curve twice at that point) and take the third point on the intersection of the line and the curve, and reflect once again. For example, continuing the last example, if $P = (12, 36)$, then we can compute the tangent line to the curve $y^2 = x^3 - 36x$ as follows. We can use implicit differentiation to compute that $2y \cdot y' = 3x^2 - 36$, and so $y' = (3x^2 - 36)/2y$. At P, this gives us that the tangent line has slope $(3 \cdot 144 - 36)/72 = 11/2$, and so we draw a picture with the curve, the base point P, the tangent line (in point-slope form) $y - 36 = 11/2 \cdot (x - 12)$, the third intersection point with the curve $(25/4, 35/8)$, and the reflected point is $2P = (25/4, -35/8)$:



Note that this gives us a procedure for generating many further examples of rational right triangles with area 6. Here, by using the bijection above, we find that this rational point exhibits the rational right triangle with side lengths $(7/10, 120/7, -1201/70)$ and area 6 (note: if you plug in the point $2P$ on the elliptic curve directly into the bijection above, you get three negative numbers; above, we didn't worry about the positivity of solutions required by side lengths of triangles, but this is easy to fix as the degree of each term in $x^2+y^2-z^2$ and $xy/2$ is 2). We can generate many more, in fact infinitely many more, such rational triangles by repeating this and continually doubling the points we obtain. This gives the rational points on the curve:

$$4P = \left(\frac{1442401}{19600}, \frac{1726556399}{2744000}\right),$$

$$8P = \left(\frac{4386303618090112563849601}{23371016471594322055840 0}, \frac{87043691090855808282759356506262544 01}{11298385851246361973721668449644800 0}\right)$$

$$16P = \left(\frac{2113^2 \cdot 33087169^2 \cdot 30331895721797713497743460237487159603 3^2}{2^8 \cdot 5^2 \cdot 7^2 \cdot 31^2 \cdot 1151^2 \cdot 1201^2 \cdot 1249^2 \cdot 10177^2 \cdot 106207^2 \cdot 730753^2 \cdot 205792513^2 \cdot 1727438169601^2},\right.$$
$$(2113 \cdot 16127 \cdot 33087169 \cdot 1807396543 \cdot 1836702719 \cdot 8904449709313 \cdot 12811537941060031$$
$$\times \cdot 18850266152574792457729 \cdot 1277224456920946252154497$$
$$\times 30331895721797713497743460237487159603 3/(2^{12} \cdot 5^3 \cdot 7^3 \cdot 31^3 \cdot 1151^3 \cdot 1201^3 \cdot 1249^3$$
$$\left.\times 10177^3 \cdot 106207^3 \cdot 730753^3 \cdot 205792513^3 \cdot 1727438169601^3)\right)$$

The numerator of the second fraction, in reduced terms has 148 digits! It would certainly be infeasible to find these points, and the corresponding right triangles, by "brute force." The equations for these fractions get extremely complicated quickly (which is related to why adding points on elliptic curves is handy for cryptography). Finally, let"s note that the group operation of "adding" points on an elliptic curve, although is natural from a geometric/graphical perspective, can be defined entirely algebraically. That is, the equation for a line between two points can be written down easily, and finding the third point of intersection can be done by solving the equation for the line for one of the variables and plugging into the other equation which gives a solvable cubic equation

(this cubic polynomial already has two known roots, facilitating its factorization). Reflecting a point is also easy algebraically; one just negates the y-variable. Thus, one can write down explicit equations algebraically defining the elliptic curve point addition property. These equations will not look pretty, but they serve an important purpose. Namely, you can study elliptic curves and their group laws in more general contexts. For instance, if the coefficients A, B are rational numbers (and, technically, if there is a rational solution to the equation $y^2 = x^3 + Ax + B$), then we can define a rational elliptic curve as the set of rational solutions to the same equation, satisfying the same equations for adding points. These equations preserve rationality and make sense even though you can't draw an equation for a line or a curve over the rationals (which have loads of gaps at irrational numbers) in the same way that you can draw the pictures over the real numbers.

As we shall see, elliptic curve cryptography also involves the elliptic curve groups $E(\mathbb{Z}_p)$, where $\mathbb{Z}_p$ is the field of p elements, consisting of the congruence classes modulo p.

4. Elliptic Curves over finite fields

The key fields we will use for elliptic curve cryptography are $\mathbb{Z}_p$. Other fields of finite order, which it turns out always have prime power order, are also important in the study of elliptic curves, but since these take more work to describe, we'll stick with prime order as its enough for our purposes.

The group of points over a finite fields a finite set. For instance, if we take the elliptic curve $E\colon y^2 = x^3 - 7x + 10$, then its discriminant is $4A^3 + 27B^2 = 4(-7)^3 + 27 \cdot 10^2 = 1328 = 2^4 \cdot 83$. If we consider the sets of solutions $E(\mathbb{Z}_p)$, this will **not** always be an elliptic curve. This is because discriminant may become zero in $\mathbb{Z}_p$, that is, if p divides the discriminant. In this case, the primes 2 and 83 are called **primes of bad reduction for** E, and all other primes not dividing the discriminant are primes of **good reduction**. Notes that the discriminant is an integer with finitely many prime factors, so there are always finitely many primes of bad reduction. If we pick a prime p of good reduction, then the discriminant modulo p is non-zero, and $E(\mathbb{Z}_p)$ really is a group. We can easily write all its points with a finite check. For instance, if we pick $p = 5$, then we want to find all $x, y \pmod 5$ solving:

$$y^2 \equiv x^3 - 7x + 10 \pmod 5.$$

For $x \equiv 0, 1, 2, 3, 4 \pmod 5$, we have the possible right hand sides of the equation:

$$10 \equiv 0, 4, 4, 16 \equiv 1, 46 \equiv 1.$$

We have studied the equation $y^2 \equiv 1 \pmod p$ for a prime p before, and there we saw that the number of solutions to this equation, or any quadratic equation, is at most 2. Thus, for each of the possible right hand sides a, if there is a square root b, then the square roots are $\pm b \pmod p$. That is, there is either no solution to the equation $y^2 \equiv a \pmod p$, there is one solution exactly if $a \equiv -a \equiv 0$ if and only if $a \equiv 0 \pmod p$, and otherwise there are two distinct roots $\pm b \pmod p$. There are good ways to determine which case you're in theoretically, but here we'll just do it by brute force for small primes. In our case, the equations we need to solve for the different values of x are:

$$y^2 \equiv 0, y^2 \equiv 4, y^2 \equiv 4, y^2 \equiv 1.$$

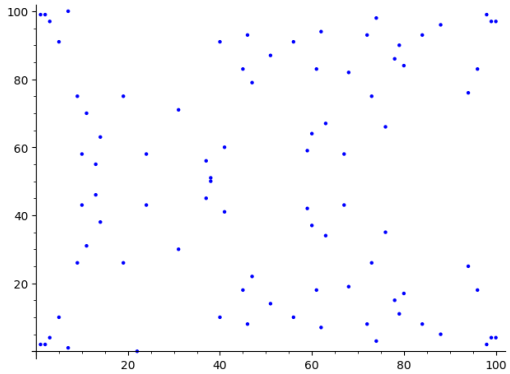
These are all perfect squares, so we can directly write down the set of all points on $E(\mathbb{Z}_5)$ (there's always a point at infinity, which we call 0):

$$E(\mathbb{Z}_5) = \{(0,0), (1,2), (1,3), (2,2), (2,3), (3,1), (3,4), (4,1), (4,4), 0\}.$$

Thus, there are 10 points on this curve. Clearly, the number of points is at most $|\mathbb{Z}_5 \times \mathbb{Z}_5| = 25$, but in fact its much smaller. In general, there is a very important theorem in elliptic curves, called the **Hasse bound**, which states that

$$||E(\mathbb{Z}_p)| - (p+1)| \leq 2\sqrt{p}.$$

That is, the number of points on E modulo p is approximately $p+1$, with error term bounded by $2\sqrt{p}$. In this example, $p+1 = 6$ and $2\sqrt{5} \approx 4.5$, so the Hasse bound says that the number of points is in $\{2, 3, 4, 5, 6, 7, 8, 9, 10\}$, and we found that the actual number is 10, just barely satisfying the bound. Writing a simple computer script, I found the points on $E(\mathbb{Z}_{101})$. There are 88 such points, and the 87 points other than infinity are displayed in the following picture:



The symmetry as $y \mapsto -y$ is visible in this picture as a reflection across the midpoint $y = 101/2$.

How do we add points in this picture? It is not as clear geometrically. We could also talk about what counts as a valid line in $\mathbb{Z}_p^2$, but instead, we'll first write down an algebraic formulation for adding real points on elliptic curves. By writing the equations of lines above and plugging them into the cubic equation (I'll skip this brute force algebra check), we find the **addition formula:** for the addition $P + Q = R$:

$$(x_P, y_P) + (x_Q, y_Q) = (x_R, y_R),$$

where

$$x_R = \lambda^2 - x_P - x_Q, \quad y_R = \lambda(x_P - x_R) - y_P, \quad \lambda = \frac{y_Q - y_P}{x_Q - x_P}. \tag{4.1}$$

Note that this doesn't make sense if the two points have the same x-coordinate, but since if they have the same x-coordinate they have plus or minus the same y-coordinate, this is either a point doubling (see below), or we're adding a point to its inverse and we know its sum is 0 anyways. The point is, these equations make sense as long as you can add, subtract, multiply, and divide, that is, as long as we are working over a **field**. For point doubling, where one has to use a tangent line, we can use the same equation, but we have to replace λ by

$$\lambda = \frac{3x_P^2 + A}{2y_P}, \tag{4.2}$$

where the elliptic curve is $y^2 = x^3 + Ax + B$.

For example, with our curve $y^2 \equiv x^3 - 7x + 10 \pmod 5$, we can add our points $P = (1, 2)$ and $Q = (4, 1)$ by following the equations:

$$\lambda = \frac{1-2}{4-1} \equiv -3^{-1} \equiv -2 \equiv 3 \pmod 5, x_R \equiv 9-1-4 \equiv 4 \pmod 5, y_R \equiv 3(1-4)-2 \equiv -11 \equiv 4 \pmod 5.$$

Thus, we have

$$P + Q = (4, 4),$$

which was on our list of points. Let's try the doubling formula in an example. If $P = (1, 2)$, and we double, then we compute

$$\lambda = \frac{3 \cdot 1^2 - 7}{2 \cdot 2} \equiv -4 \cdot 4^{-1} \equiv -1 \pmod 5.$$

Thus, we find

$$x_{2P} \equiv (-1)^2 - 1 - 1 \equiv 4 \pmod 5, y_{2P} \equiv -(1-4) - 2 \equiv 3 - 2 \equiv 1 \pmod 5,$$

and so

$$2P = (4, 1).$$

5. The discrete logarithm problem

The structure of elliptic curves can be used in several ways to build interesting cyrptoschemes. The key one we will consider is based on the *discrete log problem*, which is as follows.

Problem 1 (Discrete Log Problem (DLP))**.** *Let G be a finite cyclic group. That is, G is finite and it is generated by an element $a \in G$:*

$$\langle a \rangle = \{1, a, a^2, \ldots\} = G.$$

Given $b \in G$, find the smallest nonnegative integer x such that

$$a^x = b.$$

We call x the discrete logarithm of b with respect to base a and denote it by $x = \log_a(b)$.

Example. As we saw before, the groups $\mathbb{Z}_p^*$ for primes p are always cyclic. Recall that in this case, a generator for $\mathbb{Z}_p^*$ is called a primitive root. If we pick $p = 101$, then it turns out that 2 is a primitive root. This means that

$$\langle 2 \rangle = \mathbb{Z}_{101}^*,$$

or equivalently, $\text{ord}(2) = 100 = \varphi(101) = |\mathbb{Z}_{101}^*|$, or equivalently still, $\log_2(1) = 100$. If you wanted to check this by brute force, we could first note that by Fermat's Little Theorem and the remarks in its proof, we have that the order of 2 divides 100; then we can check that for the 9 divisors of 100, $1, 2, 4, 5, 10, 20, 25, 50, 100$, none of those powers of 2 are $1 \pmod{101}$.

Thus, every non-zero congruence class modulo 101 is a power of 2. The discrete logarithm problem asks us to find, given a particular congruence class, this power. As an example, we could find the discrete log of 3 with respect to this choice of generator. It turns out that

$$\log_2(3) = 69.$$

Indeed, we have that $2^{100} \equiv 1 \pmod{101}$, and 2^{50} is a square root of 1 modulo 101, so therefore $2^{50} \equiv -1 \pmod{101}$. Thus, we have

$$2^{69} \equiv 2^{50} \cdot 2^{19} \equiv -(2^7)^2 \cdot 32 \equiv -27^2 \cdot 32 \equiv -27 \cdot 864 \equiv -27 \cdot 56 \equiv -1512 \equiv -98 \equiv 3 \pmod{101}.$$

We also can't have a smaller power of 2 giving $3 \pmod{101}$, since if $2^x \equiv 3 \pmod{101}$, $0 < x < 69$, then

$$2^{69-x} \equiv 1 \pmod{101},$$

but we know this is impossible since the order of 2 is 100 by choice.

As we can already start to see, the discrete logarithm is not easy to compute. The difficulty of computing this is the basis for the security of the cryptoscheme we'll build.

6. ElGamal cyrptoscheme

There are a number of ways to use the group structure of elliptic curves to build cryptoschemes. We will focus on one example here; the **ElGamal** system. The ElGamal system works for any choice of a finite cyclic group. The classical situation is for the group $\mathbb{Z}_p^*$ for a prime p, so we'll first describe it in this case.

To start, pick a prime p, and a generator α of $\mathbb{Z}_p^*$ (that is, an element of order $p-1$). In practice, p may be a prime with 200-300 digits, say. Then choose a (possibly random) integer d with

$$2 \leq d \leq p-2.$$

To set up the scheme, Alice computes

$$\beta \equiv \alpha^d \pmod{p}.$$

The **public key** is then

$$(p, \alpha, \beta).$$

Alice's **private key** is

$$d.$$

Encryption works as follows. Take a message m, which here is an integer $0 \leq m < p$. Bob chooses a random integer $1 \leq k < p$ (which he keeps a secret). Then Bob transmits the pair

$$(r, t) := (\alpha^k \pmod{p}, (m \cdot \beta^k) \pmod{p}).$$

Alice can decrypt the message by computing $tr^{-d} \pmod{p}$. This works since

$$tr^{-d} \equiv \beta^k m (\alpha^k)^{-d} \equiv m \cdot \alpha^{kd} \cdot \alpha^{-kd} \equiv m \pmod{p}.$$

If Eve intercepts the message, she cannot decode this as she doesn't know d, which is precisely computing a discrete logarithm. If the prime is large and the power d is chosen reasonably, this is impractical to do, making this a secure system. However, it should be noted that while it is expected, it is *not actually known*, provably, whether breaking ElGamal is equivalent to evaluating discrete logarithms.


Bob should be careful to choose a different integer k every time he sends a message, or else this won't be secure. In particular, if m is a longer message and needs to be broken into blocks, then Bob should choose a different k on each block.

For instance, if he encrypts two messages (or two blocks) m_1 and m_2 using the same k, then he produces the ciphertexts

$$(r_1, t_1) = (\alpha^k, \beta^k m_1),$$
$$(r_2, t_2) = (\alpha^k, \beta^k m_2).$$

But, we have $t_2 t_1^{-1} \equiv m_2 m_1^{-1} \pmod{p}$, and in particular, $m_2 \equiv t_2 t_1^{-1} m_1 \pmod{p}$. Thus, if Eve can discover the plaintext for m_1, then she can immediately compute the other plaintext m_2.

Example. Suppose Alice chooses $p = 107$, $\alpha = 2$, and $d = 67$. Then she computes that

$$\beta = 2^{67} \equiv 94 \pmod{107}.$$

Her public key is $(107, 2, 94)$, and her private key is 67. Suppose Bob uses ASCII to encode characters, and wants to send the letter "B," which in ASCII is 66. He randomly chooses $k = 45$ and encrypts $m = 66$ as

$$(r, t) = (\alpha^k, \beta^k m) \equiv (2^{45}, 94^{45} \cdot 66) \equiv (28, 9) \pmod{107}.$$

He sends the ciphertext $(28, 9)$ to Alice.

When Alice receives this, she can decrypt by computing

$$tr^{-d} \equiv 9 \cdot 28^{-67} \equiv 9 \cdot 28^{106-67} \equiv 9 \cdot 28^{39} \equiv 9 \cdot 43 \equiv 66 \pmod{107}.$$

7. Elliptic Curve ElGamal

We can modify the above to work for elliptic curves over finite fields. We first require a little background. The definition of orders in groups of points on elliptic curves is just like the definition of orders in $\mathbb{Z}_p^*$.

Namely, the order of a point P, denoted $\mathrm{ord}(P)$, is the smallest positive multiple of P (instead of power, since the group operation is written as $+$, not as times) which gives the identity point 0 (the point at infinity). This exists because we are dealing with a finite group if we're working over a field, and by Lagrange's Theorem which we saw for general groups before, the order is a divisor of the order of the group, $|E(\mathbb{Z}_p)|$. Also just as for $\mathbb{Z}_p$, if we take a point $P \in E(\mathbb{Z}_p)$, then we have the cyclic subgroup generated by P, given by

$$\langle P \rangle = \{0, P, 2P, \ldots, (\mathrm{ord}(P) - 1)P\}.$$

To setup ElGamal for elliptic curves, Alice picks an elliptic curve E and a prime p of good reduction for E (that is, which doesn't divide the discriminant of E).Then she picks a point Q_a on $E(\mathbb{Z}_p)$ which has a large order. Ideally, this order should be chosen to be a large prime which is approximately the number of points on $E(\mathbb{Z}_p)$, but this is not always easy to do.

A quick note: above we chose $\alpha \in \mathbb{Z}_p^*$ to be a primitive root, that is, a generator of the group. That wasn't strictly speaking necessary. The cyclic subgroup generated by α is always cyclic (generated by one element) by choice, so still can be used in ElGamal. However, what we want is the order of α to be large so that the discrete log problem is hard to solve. In the case of $\mathbb{Z}_p^*$, it is always possible to pick the order to be the size of the group, so that's what we do. For elliptic curves, however, we will have to be more flexible.

To search for a Q_a of large order, one can try points mod p and take very large multiples of them such as $(m!)Q_a$; if this is not 0, then the order is at least $m + 1$. Alternatively, Alice could choose

an elliptic curve and a prime such that $E(\mathbb{Z}_p)$ has a prime number of points on it, and then all points other than the point at infinity are generators of the group.

Having picked the curve E, the prime p, and the point Q_a, Alice picks a positive integer d which is less than the number of points in $E(\mathbb{Z}_p)$. Note that by the Hasse bound above, Alice is free to pick any d which is less than $p - 2\sqrt{p}$, and doesn't actually need to know $|E(\mathbb{Z}_p)|$. Alice now computes $Q_b := dQ_a$ using the point addition formula above. In practice, one may try to double the point a number of times to try to use the fewest additions possible (see the numerical example below).

Alice's **public key** is then

$$(E, p, Q_a, Q_b).$$

Her **private key** is

$$d.$$

Suppose now that Bob wishes to send a message $P = (x, y) \in E(\mathbb{Z}_p)$, a pair of numbers modulo p lying on the curve. In practice, we will pick this p to be such that $|E(\mathbb{Z}_p)|$ is large, and then encode a set of possible bit messages or string messages first by mapping them in a reversible way to points on the elliptic curve. At random, Bob chooses an integer k less than the number of points on $E(\mathbb{Z}_p)$ (again, he's free to choose any $0 \leq k < p - 2\sqrt{p}$. He then computes

$$Q_r := kQ_a, \qquad Q_t := kQ_b + P.$$

He sends the message (Q_r, Q_t) to Alice.

To decrypt, Alice receives (Q_r, Q_t) and computes

$$Q_t - dQ_r = kQ_b + P - dkQ_a = kdQ_a + P - dkQ_a = P.$$

In practice, this subtraction is performed by adding $d \cdot (-Q_r)$, where recall that additive inverse in the real plane is just a reflection across the x-axis, and so in general its just negative the y-coordinate.

Example. Suppose Alice chooses the elliptic curve $E\colon y^2 = x^3 + 7x + 1$. The discriminant of E is

$$4 \cdot 7^3 + 27 \cdot 1^2 = 1399,$$

which is prime. Thus, she may choose any prime p not equal to 1399 (so that $E(\mathbb{Z}_p)$ is an elliptic curve). Let's pick the prime $p = 44927$. We'll also pick the point

$$Q_a := (7772, 14369).$$

This is indeed a point on the curve as

$$7772^3 + 7 \cdot 7772 + 1 = 469459818053 \equiv 28596 \equiv 206468161 = 14369^2 \pmod{44927}.$$

It turns out that this has order 44651, and that we're in the special case mentioned above where $|E(\mathbb{Z}_p)| = 44651$ is prime so that all non-identity points on the curve have this order. So finding this curve and choice of prime to make this happen may not be obvious, but given it, finding points of large order is trivial.

Now Alice chooses the private key $d = 22105$. She needs to compute $222105Q_a = Q_b$. To do so, she can use a computer and successive doubling as follows. Using (4.1) (recalling that for point doubling, you must use λ as given by (4.2)), we can compute

$$2Q_a = (12718, 14083)$$

$$4Q_a = (25359, 35260),$$

and so forth until we get to

$$2^{14}Q_a = 16384Q_a = (21448, 31413).$$

This gets us "closer" to dQ_a. Now one can get the binary representation for d:

$$d = 101011001011001.$$

Reading off, this tells us that

$$dQ_a = 2^{14}Q_a + 2^{12}Q_a + 2^{10}Q_a + 2^9Q_a + 2^6Q_a + 2^4Q_a + 2^3Q_a + Q_a.$$

In computing $2^{14}Q_a$, we have already computed all of the lower powers of 2 multiples of Q_a along the way. So we computed 14 point additions to get to $2^{14}Q_a$, and then 7 more must be computed to use the last displayed equation, and so overall, we only had to compute 21 point additions to compute $22105Q_a$. Each of these is just a mod p computation, and pretty quick; unlike what we saw over $\mathbb{Q}$, the points don't get way bigger as we keep adding, they just get very jumbled up mod p. At any rate, using this or another method, we compute

$$Q_b = dQ_a = (39061, 4109).$$

In practice, this is quite fast, for instance, on my laptop, it took my computer 483 microseconds, so basically half a milisecond, to compute dQ_a. The public key is thus

$$(y^2 = x^3 + 7x + 1, 44927, (7772, 14369), (39061, 4109)).$$

Let's say we want to encode the message

$$P = (14605, 29833).$$

Bob chooses a random integer $k < p - 2\sqrt{p} \approx 44503.1$. Suppose its $k = 23207$. By a suitable method such as the "point doubling" method above, Bob computes

$$Q_r = kQ_a = (30566, 37885), \qquad Q_t = kQ_b + P = (35487, 8262) + P = (40194, 40273)$$

Alice would receive the ciphertext

$$(40194, 40273)$$

and decrypt as

$$Q_t - dQ_r = Q_t + (35487, 36665) = (14605, 29833),$$

recovering the original plaintext.

The only "reasonable" attack to try is to compute discrete logarithms. There is a way to do this, particularly the famous baby-step giant-step algorithm, but the point is this is way too slow for large primes.

As for advantages over cryptoschemes like RSA, as mentioned, it doesn't suffer some of the same attacks like common modulus (in fact, the NSA/NIST publishes an elliptic curve that it recommends for use, although there are concerns that a backdoor could potentially have been included). The other main advantage is smaller key sizes. Much research has been done into the speed of integer factorization as well as computing discrete logs on elliptic curves. Based on the known algorithms, it is estimated that an elliptic curve cryptosystem with a (typical) key size of 256 bits provides as much security as an RSA system with 3000 bits. This means that its way faster to use in practice. It is a little slower to do the point addition than modular exponentiation, but not by a lot, and the

way smaller key sizes more than makes up for it. So for many https websites, the security in the "s" is based on elliptic curves, and its essential for use on things like the small computer chips on credit cards.

Overall, computing the discrete logarithms for elliptic curves groups is also much harder than computing discrete logarithms for $\mathbb{Z}_p^*$. The methods known for computing discrete logarithms are also only effective when the number of points on the elliptic curve has only small divisors, another reason why picking a curve with a large prime number of points is a good idea.

8. Solving the discrete log problem on elliptic curves

Although there is no known fast way to compute discrete logarithms, there are some basic methods, which are fortunately impractical for the sizes of primes usually used in elliptic curve cryptography. One, which we just mentioned, is the *baby-step giant-step* algorithm. This solves the discrete log problem on $E(\mathbb{Z}_p)$ in approximately $O(\sqrt{p})$ steps. The idea is simple.

Remark. Technically, the discrete log problem as above asks for the **smallest** positive integer d solving $Q_b = dQ_a$. However, one can check directly that the decryption step above only relied on that equation being true, and not on d being minimal. So if we Eve can find **any** d making this equation true, then she can decrypt the message. Any solution will be the same as the discrete log plus a multiple of the order of Q_a, but Eve doesn't need to figure out this order to decrypt (although there are fast ways to do so without finding all the points explicitly like we did above). Thus, in the algorithm below, we just need to find a single solution by any means possible.

Baby-step Giant-step Goal: Given an elliptic curve $E(\mathbb{Z}_p)$ over a field of prime order p, and points $Q_a, Q_b \in E(\mathbb{Z}_p)$ such that $dQ_a = Q_b$, find d.

Step 1: Choose a natural number M and compute two lists:

$$\{xQ_a : 0 \leq x \leq M-1\} = \{0, Q_a, 2Q_a, 3Q_a, \dots, (M-1)Q_a\},$$
$$\{Q_b - MyQ_a : 0 \leq y \leq M-1\}.$$

Step 2: Compare the two lists. If there are any common elements, we've found a solution. That is, suppose that we find

$$xQ_a = Q_b - MyQ_a.$$

Then we have $(x + My)Q_a = Q_b$, and so $d = x + My$ is a solution to our discrete log problem.

By a pigeonhole-type argument, we expect a common point to exist on the two lists to exist when $M^2 \geq |E(\mathbb{Z}_p)| \approx p$ (which is where the square root bound came from). Basically, when M gets big enough, we will get a collision, but the larger M gets, the more time it takes to compute all these points.

Example. If we pick the same elliptic curve $E\colon y^2 = x^3 + 7x + 1$, and work modulo the prime $p = 997$, then we can consider the point $Q_a = (1, 3) \in E(\mathbb{Z}_{997})$. Then $50Q_a = (991, 133) =: Q_b$ on this curve. If we are only given Q_a and Q_b, and want to find the number 50, then we keep listing out more and more multiples xQ_a and $q_b - MyQ_a$ for increasing values of x, y until we find a collision between these two lists. It turns out that $|E(\mathbb{Z}_{997})| = 1016$, so we expect that we should find this by the time M is about $\sqrt{1016} \approx 32$. So let's try $M = 32$. The first set is:

$$\{xQ_a : 0 \leq x \leq M-1\} = \{0, (1,3), (887, 182), (640, 758), (99, 70), (201, 959), (701, 639), (619, 837),$$
$$(181, 935), (122, 13), (245, 842), (298, 459), (461, 610), (98, 361), (634, 734), (951, 990), (486, 212), (736, 852),$$

$$(568,453),(139,837),(281,819),(26,209),(738,943),(979,5),(249,233),(705,920),(239,160),(904,369),$$

$$(53,751),(324:29),(166,243),(725,666)\}.$$

The second set is:

$$\{Q_b - MyQ_a : 0 \le y \le M-1\} = \{(991,133),(568,453),(634,263),(119,556),(179,43),(736,852),(951,7),$$

$$(806,507),(753,492),(486,212),(486,785),(753,505),(806,490),(951,990),(736,145),(179,954),(119,441),$$

$$(634,734),(568,544),(991,864),(709,330),(98,361),(139,160),(4,517),(409,825),(461,610),(281,178),$$

$$(351,818),(711,908),(298,459),(26,788),(165,621)\}.$$

These two sets have several points in common, namely

$$\{xQ_a : 0 \le x \le M-1\} \cap \{Q_b - MyQ_a : 0 \le y \le M-1\} = \{(298,459),(461,610),(98,361),(634,734),(951,990),$$

$$(486,212),(736,852),(568,453)\}$$

Let's take the first one, $(298, 459)$, as we only need one. This corresponds to $x = 11$, $y = 29$. Thus, a d solving the equation $dQ_a = Q_b$ is $d = x + My = 939$. It turns out the order of Q_a on the curve is 127, so indeed

$$939Q_a = (7 \cdot 127 + 50)Q_a = 7 \cdot 0 + 50Q_a = 50Q_a = Q_b.$$

Example. Baby-step giant-step also works for other discrete log problems, such as in $\mathbb{Z}_p^*$. For instance, above, we saw $2^{69} \equiv 3 \pmod{101}$. We can write the same sets, just remembering to write multiplicatively instead of additively, and look for a coincidence. This is also expected to work by the time M is about $\sqrt{p}$, so let's try $M = 10$. Then we have (all computations done mod 101):

$$\{2^x : 0 \le x \le 9\} = \{1,2,4,8,16,32,64,27,54,7\}$$

$$\{3 \cdot 2^{-My} : 0 \le y \le 9\} = \{3,94,50,18,59,98,7,51,83,42\}.$$

The number 7 is on both of these lists, indeed, $7 \equiv 2^9 \equiv 3 \cdot 2^{-6 \cdot 10} \pmod{101}$, and so $2^{69} \equiv 3 \pmod{101}$, as desired.

There are other algorithms for solving discrete log problems as well. One of the reasons that elliptic curve cryptography is better than, say, using ElGamal over $\mathbb{Z}_p^*$, is that is is not as susceptible to algorithms using "sieving" techniques. Basically, if you multiply two small numbers mod p together, you still obtain a small number mod p. One can use this observation to try to compute a large number of relations among small primes and use those relations to compute discrete logs of many small primes to allow discrete logs to be computed quickly.

As the examples of point addition above show, the same "smallness" feature is not present for elliptic curves. Firstly, even if the x-coordinate is small, the y-coordinate is basically random-looking, and can be large. The modular divisions in the point addition law also wildly move around points, and so the sum of two "small" points is often not small. For instance, in the above baby-step giant-step example, for the point $Q_a = (1, 3)$, we had $2Q_a = (887, 182)$.

9. Symmetric elliptic curve cryptosystem: Diffie-Hellman

Although elliptic curve ElGamal is very secure, and it is much faster than RSA, in general, public-key cryptoschemes are not very fast if you need to exchange very large messages. For instance, the speed is very good for transmitting a sample of text, such as your credit card number and address, but if you need to send really gigabytes of data, even elliptic curve ElGamal is too slow in practice.

Symmetric cryptosystems don't usually require as much computation and can send large amounts of data more quickly. In practice, its usually best if a lot of information will be exchanged between two parties to have a *key exchange* first, and to use a symmetric system. Of course, the key is a small amount of data, so a very secure, but slower cryptoscheme, such as an asymmetric one like RSA or elliptic curve ElGamal can be used to exchange this key. For instance, Bob can create a public key, and then Alice can produce a key to be used in a symmetric cryptoscheme, and send it encrypted to Bob using his public key. Bob can then use his private key to decrypt and then the key exchange is complete.

Here, we'll consider another way of doing a key exchange, **Diffie-Hellman system.** As with ElGamal, it works for any finite cyclic group, so we'll first describe how it works in $\mathbb{Z}_p^*$ and then give the analogue for elliptic curves.

9.1. **Diffie-Hellman over $\mathbb{Z}_p^*$. Step 1:** Alice and Bob choose a common large prime p and a primitive root g modulo p (that is, an element of order $p-1$).

Step 2: Alice chooses a secret integer a and sends Bob $g^a \pmod p$.

Step 3: Bob chooses a secret integer b and sends Alice $g^b \pmod p$.

Step 4: The secret key is then $g^{ab} \pmod p$, which both of them are able to compute. That is, Alice has a and g^b, so she raises g^b to the a-th power, and similarly for Bob.

Note that if Eve is listening in, she will know p, g, g^a, and g^b (the last three mod p), but there is no obvious way to compute g^{ab} without either knowing a or b, which amounts to solving a discrete log problem.

9.2. **Diffie-Hellman for elliptic curves.** The analogous procedure is as follows: **Step 1:** Alice and Bob choose a common elliptic curve E over a finite field $\mathbb{Z}_p$, and a point $P \in E(\mathbb{Z}_p)$ of large order N.

Step 2: Alice chooses a secret integer $a < N$ and sends Bob $Q_a = aP$.

Step 3: Bob chooses a secret integer $b < N$ and sends Alice $Q_b = bP$.

Step 4: The secret key is then $Q_{ab} = abP$, which both of them are able to compute. That is, Alice has a and bP, so she computes $(ab)P$ as $a(bP)$.

Example. Suppose we use the elliptic curve $E\colon y^2 = x^3 + 7x + 1$, and the prime $p = 44927$. We also use the point $P = (27844, 29401)$. Suppose further that Alice chooses secret number $a = 40006$ and Bob chooses $b = 18846$. Alice then computes

$$Q_a = aP = (3454, 34367)$$

and Bob computes

$$Q_b = bP = (22472, 6971).$$

Both are able to compute the secret key

$$Q_{ab} = aQ_b = bQ_a = (2147, 22480).$$

This can now be used in a symmetric key cryptosystem. For instance, one can use sufficiently hard-to-crack Feistel ciphers (the descendants of the now-defunct DES, such as the newer AES (advanced encryption standard)).

10. Issue with Diffie-Hellman and solution

The Diffie-Hellman protocol does suffer from one issue, however, it is susceptible to "man-in-the-middle" attacks, whereby someone simultaneously impersonates Alice and Bob to each other and Alice and Bob will not be able to tell. However, this problem can be solved by adding in **authentication**, that is, by using **digital signatures.** ElGamal on elliptic curves can also be used to create digital signatures that Alice and Bob can use to authenticate each other's messages. Thus, in practice, one may combine both algorithms to set up a secure exchange.

To describe this, suppose that an adversary Mallory impersonates Alice to Bob and Bob to Alice, and does a key exchange with both of them. Then Mallory can decode messages from Alice, re-encrypt them and send them to Bob, and vice versa. Mallory can keep doing this and Alice and Bob won't know.

This can be solved by having Alice and Bob put digital signatures on their communications during the Diffie-Hellman key creation. Thus, each person knows that Mallory is not impersonating either of them.

We can use elliptic curves to create such digital signatures. We should also mention that this is very useful in general, obviously to sign documents electronically, and to verify email was sent from the claimed sender, etc. To do so, we need a way to bind messages to the creator, and that makes it hard to changes messages (so that no attacker alter the message). This is a separate goal from preventing a message from being deciphered, but nonetheless the same ideas useful for public key cryptography can be used for the creation of digital signatures.

10.1. Elliptic curve digital signatures via ElGamal.

Here we describe how to do this using elliptic curves. As above in ElGamal and Diffie-Hellman, all of these ideas could be done using $\mathbb{Z}_p^*$ instead of elliptic curves in a completely analogous way, but since the elliptic curve one is more secure and we've done this twice already, we'll just describe the elliptic curve version only.

Goal: Alice wishes to create a digital signature for communications with Bob (for example to do a Diffie-Hellman exchange).

Step 1: Alice creates an ElGamal public key: (E, p, Q_a, Q_b). Recall that she wants to do so in a way that Q_a is a point on E whose order has only large prime orders (or better, is a large prime). Also recall that $Q_b = dQ_a$ for Alice's private key d, and further suppose that $|E(\mathbb{Z}_p)| = N$.

Step 2: To sign a message m (which is an integer $0 \leq m \leq N$), Alice chooses at random a positive integer k such that $(k, N) = 1$.

Step 3: Alice computes

$$Q_r = kQ_a =: (x, y)$$

and

$$s \equiv k^{-1}(m - dx) \pmod{N}.$$

Step 4: Alice now sends Bob her signed message

$$(m, Q_r, s).$$

Step 5: Bob verifies Alice's signature by checking if

$$xQ_b + sQ_r = mQ_a.$$

If they are equal, the signature is accepted, otherwise, it is rejected.

We can check that this equation must hold if the signature is valid as:

$$xQ_b + sQ_r = x(dQ_a) + s(kQ_a) = xdQ_a + (m - dx)Q_a = mQ_a,$$

using the equations above and the fact that the order of Q_a divides N (we saw this when we discussed Lagrange's Theorem).

The security is based on the difficulty of computing a discrete logarithm and the randomness of the chosen k. It doesn't matter if the number N is known, as mentioned above, this isn't very hard to compute anyways.

Example. Alice again uses the elliptic curve $E\colon y^2 = x^3 + 7x + 1$, and prime $p = 44927$. She sets up an ElGamal public key using the point $Q_a = (3174, 1067)$ and private key $d = 25661$, giving $Q_d = dQ_a = (38921, 25436)$. The message to be sent is $m = 17781$, and the secret number is $k = 33050$. Alice first finds the number of points on the elliptic curve, $N = 44651$, which it turns out is prime. She now computes

$$Q_r = kQ_a = (11123, 34794) =: (x, y)$$

and

$$s \equiv k^{-1}(m - dx) \equiv 42665 \pmod{N}.$$

She sends the pair (Q_r, s) to Bob, who can verify the signature by checking

$$xQ_b + sQ_r = (29063, 26534) + (36219, 42811) = (35670, 7590) = mQ_a.$$

Thus, Bob accepts the signature.

11. Amuse-bouche: The Congruent Number Problem

This final section won't be part of the exam, but since it is beautiful, and wraps up the first motivating part of the elliptic curves section nicely, I thought I would include it for fun.

We saw above that n is congruent if and only if the elliptic curve $E_n\colon y^2 = x^3 - n^2x$ has a rational point in the plane with non-zero y coordinate on it. The curve has four natural points on it, $(0,0), (0, \pm n)$, and the point at infinitely. It turns out (its not a hard thing to check), these are the **only** points of finite order, also called **torsion points**. Namely, they are of order 1 (the identity element has order 1 in all groups) or of order 2 (geometrically, this is the fact that the curve has vertical tangent lines at these points, which is proven by computing the slopes via implicit differentiation). Thus, all of the above discussions on the congruent number problem can be combined into one over-arching result.

Theorem 11.1. *The number n is congruent if and only if there are infinitely many rational points on the elliptic curve E_n.*

It turns out that we don't have a known algorithm which checks whether an elliptic curve has infinitely many rational points on it or not. However, we **believe** we do. There is an "analytic" object $L(E, s)$ related to elliptic curves, which is essentially a function of $s \in \mathbb{C}$ built out of the point counts $|E(\mathbb{Z}_p)|$ over all primes p **simultaneously** (this is very closely related to the analytic objects I mentioned before that are tied to elliptic curves thanks to a theorem of Wiles who used

this connection to prove Fermat's Last Theorem, and this connection is why I claimed above that its not that difficult to compute the number of points on elliptic curves as this object can be quickly determined as it lives in a finite-dimensional vector space, so that one doesn't really have to compute what's happening at infinitely many primes). There is an easy to compute quantity attached to this function (the order of its zero at the $s = 1$), and this should determine the structure of the group of rational points on the curve, and in particular whether there are infinitely many of them or not. This is the subject of the \$1 Million **Birch and Swinnerton-Dyer Conjecture** (BSD), which thus says that the difficult problem of understanding the group of rational points on elliptic curves boils down to a simpler question about its solutions modulo p for varying primes p, which is "easier." Tunnell, using this and other known results about the "analytic object" proved the following beautiful result, which gives a very fast way to test if a number n is congruent.

Theorem (Tunnell)**.** *If n is a square-free integer, consider the four numbers*

$$A_n = \#\{(x,y,z) \in \mathbb{Z}^3 : 2x^2 + y^2 + 32z^2 = n,\ B_n = \#\{(x,y,z) \in \mathbb{Z}^3 : 2x^2 + y^2 + 8z^2 = n$$

$$C_n = \#\{(x,y,z) \in \mathbb{Z}^3 : 8x^2 + 2y^2 + 64z^2 = n,\ D_n = \#\{(x,y,z) \in \mathbb{Z}^3 : 8x^2 + 2y^2 + 16z^2 = n.$$

If n is congruent, then $2A_n = B_n$ is n is odd, and $2C_n = D_n$ is n is even. Conversely, if the Birch and Swinnerton-Dyer conjecture holds, then the converse is true.

For example, we saw that the "smallest" triangle representing 157 as a congruent number is absolutely enormous, and you can't find it by brute force computer search. Let's test whether its congruent using Tunnell's criterion. Since n is odd, we should compute A_n and B_n. We can just search for solutions on a computer. Its easy to bound the ranges on x, y, z we have to search through. For example, if $2x^2 + y^2 + 32z^2 = 157$, then $2x^2 \leq 157$ and so $|x| \leq 8$. In this case, we find no solutions to the above quadratic equations, and so $2C_n = D_n = 0$. Thus, assuming the Birch and Swinnerton-Dyer conjecture, we have that 157 is congruent, as we claimed above. We can show unconditionally that 1 isn't congruent (a famous result of Fermat) via Tunnell's criterion by noting that the only solutions to the equations defining A_n and B_n are $(x,y,z) = (0,\pm 1,0)$, so that $A_n = B_n = 2$. But $2 \cdot 2 \neq 1$, and so 1 must not be congruent. To search for solutions in these cases, I only needed to test the values of 3 triples for each of A_n and B_n, which is pretty efficient. In general, Bach and Ryan proved that for general n, this criterion can be checked in about time $O(n^{\frac{1}{2}})$. We conclude with an interesting consequence of all of the above, which is non-obvious.

Corollary 11.2. *If n is congruent, then there exist infinitely many rational right triangles with area n.*

12. Supplemental Problems

(1) Consider the elliptic curve $E\colon y^2 = x^3 + x + 1$ modulo the prime $p = 5$. It turns out that $|E(\mathbb{Z}_5)|$, the number of points on the curve, is 9. This problem concerns ElGamal for this curve. (Remember, on the exam I'll hand you the formula for point addition and point doubling, as on the sample exam).
   (a) Letting Q_a be the initial point $(0,1)$, and choosing the private key $d = 5$, compute the public key.
   (b) Encrypt the message $P = (2,1)$ using the choice $k = 2$.

(c) Decrypt the ciphertext $((2,1),(0,1)$.

(2) Let $p = 13$ and consider the generator $\alpha = 2$ of $\mathbb{Z}_{13}^*$. Consider ElGamal for this setup.
   (a) Let $d = 5$ be the private key. Use this to produce the public key.
   (b) Use your answer from (a) and the choice of integer $k = 3$ to encrypt the message 7.
   (c) Decrypt the ciphertext $(11, 12)$.

(3) With the same elliptic curve E and prime p as in problem 1 and the P is the point Q_a from (1) as well, suppose Diffie-Hellman is performed to exchange a secret key between Alice and Bob. Suppose that Alice chooses secret integer 4 and Bob chooses integer 5. Compute the shared secret key.

(4) Give the shared secret key using Diffie-Hellman using the prime $p = 13$, the generator $g = 2$, and secret integers $a = 5$, $b = 6$.

(5) Using the same ElGamal setup from problem (1), and message $m = 8$, and secret number $k = 5$, compute the signed message

$$(m, Q_r, s)$$

where $Q_r = kQ_a = (x, y)$ and $s \equiv k^{-1}(m - dx) \pmod 9$. Note: on the exam, I will give you the formula for the digital signature production or verification, if there's a problem on digital signatures.

13. Answers to Supplemental problems

Here I give mostly numerical answers to the above, both so that you can practice obtaining the answers yourself as better practice for the exam. These answers give a way to check your work, and give you confidence that you're doing it right.

(1) (a) We need to compute $Q_b = 5Q_a$. We compute $2Q_a = (4,2)$, $4Q_a = 2(4,2) = (3,4)$, $5Q_a = Q_q + 4Q_a = (3,1)$. Thus, the public key is $(y^2 = x^3 + x + 1, 5, (0,1), (3,1)$.
   (b) We have to compute $Q_r = kQ_a = (4,2)$, as listed above (so on the exam, here you could just cite the above to save computation time). We also compute $Q_t = kQ_b + P = 2(3,1) + P = (0,1) + (2,1) = (3,4)$. The ciphertext is thus $((4,2),(3,4))$.
   (c) We compute $(0,1) - 5(2,1) = (0,1) + 5(2,-1) = (0,1) - (2,4) = (0,1) + (2,1) = (3,4)$.

(2) (a) We compute $2^5 \equiv 16 \cdot 2 \equiv 6 \pmod{13}$. Thus, the public key is $(13, 2, 6)$.
   (b) The value of r in the ciphertext (r,t) is $\alpha^k \equiv 8 \pmod{13}$. The value of t is $7 \cdot 6^3 \equiv 4 \pmod{13}$. Thus, the ciphertext is $(8,4)$.
   (c) The plaintext is $12 \cdot 11^{-5} \equiv 12 \cdot (11^{-1})^5 \equiv 12 \cdot 6^5 \equiv 11 \pmod{13}$.

(3) Above, we had computed that $4P = (3,4)$ and that $5P = (3,1)$. (Recall, on the exam, you can always reuse computations you've already done on previous problems). We can either multiply the first point by 5 or the second by 4. We'll do the latter, as its just doubling a point twice. We compute $10P = 2(3,1) = (0,1)$, and then $20P = 2(0,1) = 2Q_a = (4,2)$, using the same computation from above. Note that we could take a shortcut here, as $10P = P$ **automatically** since I claimed that the size of the group is 9, and as we saw by Lagrange's theorem that therefore $9Q = 0$ for any point Q on the curve, and hence $10Q = Q$ for any point on the curve. Thus, this entire problem can be computed without over actually using the addition formula above and just using the computations already performed. Again, on an exam, such shortcuts can really help, but I will always design problems with the more straightforward approach without shortcuts being doable in the allotted time.

   In short, the shared secret key is $(4,2)$.

(4) Again, we already computed $g^a \equiv 2^5 \equiv 6 \pmod{13}$ (not that it was very difficult in this instance). Now we need $g^{ab} \equiv (g^a)^b \equiv 6^6 \equiv 10^3 \equiv 12 \pmod{13}$, the shared secret key.

(5) We have $Q_r = 5Q_a = (3,1)$, as this was already computed above. Thus, $x = 3$, and we still have $d = 5$. We then need

$$s \equiv 5^{-1}(8 - 5 \cdot 3) \equiv 2 \cdot (-7) \equiv -14 \equiv 4 \pmod{9}.$$

   Thus, the signed message is

$$(8, (3,1), 4).$$

   The problem doesn't ask for this, but just to double check, the verification is that

$$xQ_b + sQ_r = 3Q_b + 4(3,1) = 7(3,1) = 8Q_a = (0,4),$$

   which does work out.