MA 341: Notes on Finite Fields

1 The Ring $\mathbb{Z}_p[x]$

Let p be a prime. Recall that we write $\mathbb{Z}_p[x]$ for the ring of polynomials with coefficients in $\mathbb{Z}_p$. Concretely, the addition and multiplication is given by the usual operations, but coefficients are considered mod p. We have some example calculations:

$$p = 3 : (x^2 + 2x + 2) + (2x^3 + x^2 + x) = 2x^3 + 2x^2 + 3x + 2 \equiv 2x^3 + 2x^2 + 2 \bmod 3$$

$$p = 5 : (4x^2 + 2) \cdot (3x + 1) = 12x^3 + 4x^2 + 6x + 2 \equiv 2x^3 + 4x^2 + x + 2 \bmod 5.$$

So we would say that $(x^2 + 2x + 2) + (2x^3 + x^2 + x) = 2x^3 + 2x^2 + 2$ in the ring $\mathbb{Z}_3[x]$ and $(4x^2 + 2) \cdot (3x + 1) = 2x^3 + 4x^2 + x + 2$ in the ring $\mathbb{Z}_5[x]$. This is analogous to saying that $3 + 4 = 2$ in the ring $\mathbb{Z}_5$. (Equality in $\mathbb{Z}_5$ means congruence in $\mathbb{Z}$.)

The ring $\mathbb{Z}_p[x]$ has many features in common with $\mathbb{Z}$. In particular, it has a division algorithm, and hence a unique factorization theorem.

Theorem 1. *Let $f(x), g(x) \in \mathbb{Z}_p[x]$ with $g(x) \neq 0$ (the zero polynomial). Then there exist $q(x), r(x) \in \mathbb{Z}_p[x]$ with $\deg(r(x)) < \deg(g(x))$ or $r(x) = 0$ such that*

$$f(x) = g(x)q(x) + r(x).$$

Proof. Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ and $g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_0$, with $a_n, b_m \neq 0$ and $a_i, b_j \in \mathbb{Z}_p$. If $\deg(g(x)) > \deg(f(x))$, we can take $q(x) = 0$ and $r(x) = f(x)$.

Otherwise, let $q_1(x) = (a_n \cdot b_m^{-1})x^{m-n}$, where $a_n \cdot b_m^{-1}$ is computed in $\mathbb{Z}_p$. This is possible since $b_m \neq 0$, so it must be a unit, i.e. have an inverse in $\mathbb{Z}_p$. Then

$$f(x) - q(x)g(x) = (a_n x^n + \ldots a_0) - (a_n \cdot b_m^{-1} x^{n-m})(b_m x^m + \cdots + b_0),$$

and we see the leading terms cancel. Iterate this process replacing $f(x)$ with $f(x) - q(x)g(x)$ until the degree drops below that of $g(x)$, or you get the zero polynomial. □

(Small remark: we usually don't assign a degree to the zero polynomial which is why it's considered a separate case above.)

In practice, this is just performing polynomial long division, but with $\mathbb{Z}_p$ coefficients. It still works! Let's try out an example in $\mathbb{Z}_3[x]$ (so all coefficient arithmetic happens mod 3):

$$\begin{array}{r} x^2+2x \\ x+1\overline{)\,x^3+2x+1} \\ \underline{-(x^3+x^2)} \\ 2x^2+2x+1 \\ \underline{-\,(2x^2+2x)} \\ 1 \end{array}$$

So we get that $x^3+2x+1=(x^2+2x)(x+1)+1$ in $\mathbb{Z}_3[x]$. If we multiply out the right, we get $x^3+x^2+2x^2+2x+1$, whose coefficients are congruent mod 3 to those of x^3+2x+1 after collecting like terms.

As we've seen a few times now, once we have a division algorithm, we can go through the same arguments as before to get a unique factorization theorem. In this ring, the "primes" should be polynomials $f(x)\in\mathbb{Z}_p[x]$ such that $f(x)=g(x)h(x)$ implies $g(x)$ or $h(x)$ is a unit. By degree considerations, it's easy to see the units in this ring are exactly the nonzero constant polynomials (i.e. $u(x)\cdot v(x)=1$ implies $\deg(u(x))=0$). So saying $f(x)$ is a "prime" in this ring is usually referred to as saying $f(x)$ is *irreducible*.

Theorem 2. *Let $f(x)\in\mathbb{Z}_p[x]$ be nonzero. Then $f(x)$ admits a unique factorization into irreducible polynomials:*

$$f(x)=u(x)q_1(x)\cdots q_m(x)$$

with $u(x)$ a constant polynomial and $q_i(x)$ irreducible.

By unique, we mean if $v(x)q_1'(x)\cdots q_n'(x)$ is another such factorization, then $n=m$ and up to reordering $q_i'(x)=u_i(x)q_i(x)$ for some $u_i(x)$ a constant (i.e. $q_i'(x)$ is an associate of $q_i(x)$).

The proof is very similar to the others we've seen, so we'll skip it for now. But let's see how some of the computations we're used to work in $\mathbb{Z}_p[x]$.

Example. Suppose we want to solve the LDE $a(x)(x^3+1)+b(x)(x^2+1)=1$ in $\mathbb{Z}_7[x]$. Then we do the Euclidean algorithm using long division like above, computing coefficients mod 7:

$$\begin{aligned} x^3+1&=x(x^2+1)+(6x+1)\\ x^2+1&=(6x+6)(6x+1)+2. \end{aligned}$$

We stop when the remainder is a unit (i.e. the constant polynomial 2). Then we back substitute:

$$\begin{aligned} 2&=(x^2+1)-(6x+6)(6x+1)\\ &=(x^2+1)-(6x+6)((x^3+1)-x(x^2+1))\\ &=(1+x(6x+6))(x^2+1)-(6x+6)(x^3+1)\\ &=(6x^2+6x+1)(x^2+1)+(x+1)(x^3+1). \end{aligned}$$

Multiply through by $2^{-1}\equiv 4 \bmod 7$ to get

$$1=(3x^2+3x+4)(x^2+1)+(4x+4)(x^3+1)$$

so $a(x)=4x+4$ and $b(x)=3x^2+3x+4$ work.

Example. Let's look at $x^3 + 10 \in \mathbb{Z}_{11}[x]$. Then $x^3 + 10 \equiv x^3 - 1 \bmod 11$, and this factors via the difference of cubes as $x^3 - 1 = (x-1)(x^2+x+1)$. Now if x^2+x+1 factors into a product of nonunits $a(x) \cdot b(x)$, then we must have $\deg(a(x)) = 1$ and $\deg(b(x)) = 1$. But if a polynomial $f(x)$ has a linear factor, then it must have a root. (Check this! If the factor looks like $ax+b$, then the root is $-b/a \bmod 11$.)

Looking at x^2+x+1, we know it has a root if and only if the discriminant is a perfect square. Its discriminant is -3, and $\left(\frac{-3}{11}\right) = -1$ tells us that the polynomial has no root. Hence x^2+x+1 is irreducible, and $x^3+10 = (x+10)(x^2+x+1)$ is the unique factorization of x^3+10 into irreducibles in $\mathbb{Z}_{11}[x]$.

The ring $\mathbb{Z}_p[x]$ also has a very important similarity to $\mathbb{Z}$ when it comes to studying congruences. We can study congruences between elements of $\mathbb{Z}_p[x]$ in the same way as for $\mathbb{Z}$: $a(x) \equiv b(x) \bmod f(x)$ if $f(x) \mid (a(x) - b(x))$, i.e. if $f(x) \cdot g(x) = a(x) - b(x)$ for some $g(x) \in \mathbb{Z}_p[x]$.

Example. Using the division algorithm in $\mathbb{Z}_5[x]$, we get that $x^4 + 3x + 2 \equiv 3x + 1 \bmod (x^2+3)$, as $x^4 + 3x + 2 = (x^2+2)(x^2+3) + (3x+1)$. (All coefficients work mod 5 here!)

Example. Using the division algorithm in $\mathbb{Z}_3[x]$, we get that $x^7 + 2 \equiv 2x^2 + 1 \bmod (x^3+x+1)$, as $x^7 + 2 = (x^4 + 2x^2 + 2x + 1)(x^3 + x + 1) + (2x^2 + 1)$. (All coefficients work mod 3 here!)

In this way, we can start to do algebra in $\mathbb{Z}_p[x]$ mod $f(x)$ for some polynomial $f(x)$. This may seem a little strange as the coefficients of elements in $\mathbb{Z}_p[x]$ already have a "congruence flavor" to them, but it turns out this type of construction is very, very useful for many applications.

So we define the ring $\mathbb{Z}_p[x]_{f(x)}$ analogous to the way we defined $\mathbb{Z}_m$: it's the set of polynomials with $\mathbb{Z}_p$ coefficients identified when they are congruent mod $f(x)$. Before, we would think of the elements of $\mathbb{Z}_m$ as $\{0, 1, \ldots, m-1\}$ since everything in $\mathbb{Z}$ was congruent to exactly one element of this nice set. Similarly, we can write elements of $\mathbb{Z}_p[x]_{f(x)}$ in a "nice" way.

Example. Consider $f(x) = x^2 + x + 1 \in \mathbb{Z}_2[x]$. Then if $a(x) \in \mathbb{Z}_2[x]$, we can always divide $a(x)$ by x^2+x+1 to get something with remainder in the form $ax + b$ for $a, b \in \mathbb{Z}_2$. But also, if $ax + b \equiv cx + d \bmod f(x)$, then $f(x) \mid ((a-c)x + (b-d)$ would imply $(a-c)x + (b-d)$ is the zero polynomial (since all other multiples of $f(x)$ have degree at least 2). Hence $ax + b = cx + d$ in this case, and so every choice of coefficient from $\mathbb{Z}_2$ gives a new element.

In other words, $\mathbb{Z}_2[x]_{x^2+x+1} = \{0, 1, x, x+1\}$.

In fact, we have the following generalization. (Compare with the analogous theorem in $\mathbb{Z}$: every element of $\mathbb{Z}_m$ is congruent to exactly one integer between 0 and $m-1$, and two such integers are never congruent mod m.)

Proposition 3. *Let $f(x) \in \mathbb{Z}_p[x]$ with degree d. Then $|\mathbb{Z}_p[x]_f(x)| = p^d$. In fact, every element of $\mathbb{Z}_p[x]_{f(x)}$ is congruent to exactly one polynomial of degree less than d (or the zero polynomial), and two distinct such polynomials are never congruent mod $f(x)$.*

Proof. Let $g(x) \in \mathbb{Z}_p[x]$. Then write $g(x) = q(x)f(x) + r(x)$, with $\deg(r(x)) < \deg(f(x)) = d$ or $r(x) = 0$. Then $f(x) \mid (g(x) - r(x))$, so $g(x) \equiv r(x) \bmod f(x)$, and $r(x)$ has degree less than d by construction.

Now suppose $a(x), b(x) \in \mathbb{Z}_p[x]$ with $\deg(a(x)), \deg(b(x)) < d$ and $a(x) \equiv b(x) \bmod f(x)$. Then $\deg(a(x) - b(x)) < d$ but $a(x) - b(x)$ is a multiple of $f(x)$. Since every nonzero multiple of $f(x)$ has degree at least d, we see that $a(x) - b(x) = 0$, i.e. $a(x) = b(x)$. $\square$

Example. A similar line of reasoning as above shows that

$$\mathbb{Z}_3[x]_{x^2+1} = \{0, 1, 2, x, x+1, x+2, 2x, 2x+1, 2x+2\}.$$

Doing arithmetic in this ring is actually quite easy. Here, we have the relation $x^2+1 \equiv 0 \bmod (x^2+1)$, so $x^2 \equiv -1 \bmod (x^2+1)$. So to multiply $(ax+b)$ by $(cx+d)$, we just replace any instance of x^2 with -1: $(ax+b)(cx+d) = (ad+bc)x + (bd-ac)$. In this way, x in this ring acts a lot like i does in the usual complex numbers. We'll return to this idea in a little bit.

2 Finite Fields

The rings $\mathbb{Z}_p[x]_{q(x)}$ when $q(x)$ is *irreducible* have a very special structure similar to that of $\mathbb{Z}_p$ itself. They provide examples of *finite fields*.

Recall that a ring R is a set with addition and multiplication satisfying the usual properties like commutativity, associativity, the distributive law, etc. The definition of a unit makes sense in any ring, i.e. $a \in R$ is a unit if there exists $b \in R$ such that $a \cdot b = 1$. This abstracts the notion of being able to "divide" by a. A ring in which *every* nonzero element is a unit is called a **field**. You should think of these as rings where it's always legal to divide by nonzero elements.

Example. $\mathbb{Q}$ is a field, as every nonzero element has a multiplicative inverse (existence of $a^{-1} = 1/a$ as a fraction means $b = 1/a$ works for $a \cdot b = 1$).

Example. $\mathbb{R}$ is a field, as is $\mathbb{C}$, as division makes sense for every nonzero element.

Example. $\mathbb{Z}_p$ is a field for p prime, since every nonzero element is a unit.

A field which has finitely many elements is called a **finite field**. So $\mathbb{Z}_p$ for p prime gives a first example. Compare the following proposition and proof with the analogous statement that all nonzero elements of $\mathbb{Z}_p$ are units.

Proposition 4. *The ring $\mathbb{Z}_p[x]_{q(x)}$ for $q(x)$ irreducible is a finite field.*

Proof. We showed above the set is finite, so we need to show every nonzero element is a unit. Let $a(x) \in \mathbb{Z}_p[x]_{q(x)}$ be nonzero. Then since $q(x) \nmid a(x)$, we can perform the Euclidean algorithm in $\mathbb{Z}_p[x]$ with $q(x)$ and $a(x)$ to solve

$$a(x)f(x) + q(x)g(x) = 1$$

for some $f(x), g(x) \in \mathbb{Z}_p[x]$. Reducing this mod $q(x)$ gives

$$a(x) \cdot f(x) \equiv 1 \bmod q(x)$$

so $a(x)$ is a unit. □

Just as with $\mathbb{Z}_p$, the proof demonstrates exactly how to find $a(x)^{-1}$ in $\mathbb{Z}_p[x]_{q(x)}$: use the Euclidean algorithm the same way you would compute for example $5^{-1} \bmod 101$.

Example. Suppose we want to find the inverse of $x + 1$ in $\mathbb{Z}_5[x]_{x^2+2}$. We do the Euclidean algorithm with $x^2 + 2$ and $x + 1$ (with mod 5 coefficients!) to get $1 = 2(x^2 + 2) + (3x + 2)(x + 1)$ in $\mathbb{Z}_5[x]$. Then reducing mod $x^2 + 2$ gives $1 \equiv (3x + 2)(x + 1)$, so $(x + 1)^{-1} = 3x + 2$ in $\mathbb{Z}_5[x]_{x^2+2}$.

Finite fields arise in other ways too. If $\pi \in \mathbb{Z}[i]$ is a Gaussian prime, then $\mathbb{Z}[i]_\pi$ will be a finite field as well (you can prove this as an exercise).

Example. Let's look at $\mathbb{Z}[i]_3$. Note that $a+bi \equiv c+di \bmod 3$ if and only if $3 \mid (a-c)+(b-d)i$, i.e. $a \equiv c \bmod 3$ and $b \equiv d \bmod 3$. So we can write $\mathbb{Z}[i]_3 = \{0, 1, 2, i, i+1, i+2, 2i, 2i+1, 2i+2\}$ using all possible coefficients of $a + bi \bmod 3$. We add and multiply the same as with complex numbers, but always reduce real and imaginary parts mod 3: e.g. $(1 + 2i)(2 + i) \equiv 2 + 4i + i - 2 \equiv 2i \bmod 3$.

The last example might give you a feeling of deja vu: the computations feel a lot like those in the field $\mathbb{Z}_3[x]_{x^2+1}$ that we saw above. In fact, if we replace i with x, all additions and multiplications basically behave "the same."

There is a way to make this mathematically precise, which we won't go into detail here. We would say that the fields $\mathbb{Z}[i]_3$ and $\mathbb{Z}_3[x]_{x^2+1}$ are *isomorphic*. That's a fancy name to capture the idea that the arithmetic in each field feels exactly the same under some kind of substitution like $i \leftrightarrow x$.

The important fact for us is that *every finite field is "isomorphic" to $\mathbb{Z}_p[x]_{q(x)}$ for some p prime and $q(x) \in \mathbb{Z}_p[x]$ irreducible.*[1] This means if we want to understand the arithmetic of finite fields in general, it's enough to just study $\mathbb{Z}_p[x]_{q(x)}$. So let's do that a bit. Going forward, we will talk about finite fields as if they're given as $\mathbb{Z}_p[x]_{q(x)}$. Whenever we write $\mathbb{Z}_p[x]_{q(x)}$, we will assume $q(x)$ is irreducible so that this is a finite field.

Finite fields in general have a lot of similarities with $\mathbb{Z}_p$, the special case we've studied quite a bit already. For example, we have the following analogues of Fermat's Little Theorem and the Primitive Root Theorem.

Theorem 5. *Let $a(x) \in \mathbb{Z}_p[x]_{q(x)}$ be nonzero, $\deg(q(x)) = d$. Then $(a(x))^{p^d-1} \equiv 1 \bmod q(x)$. Equivalently, $(a(x))^{p^d} \equiv a(x) \bmod q(x)$ for all $a(x)$.*

You will essentially prove this on the homework, but the proof is very similar to the one for $\mathbb{Z}_p$. It really only uses the fact that the set of units is finite, and everything nonzero is a unit.

Theorem 6. *The set of units in $\mathbb{Z}_p[x]_{q(x)}$ has a generator $g(x)$ in the following sense: every nonzero $a(x) \in \mathbb{Z}_p[x]_{q(x)}$ can be written as*

$$a(x) \equiv g(x)^k \bmod q(x)$$

for some $k \geq 1$.

The proof of this theorem is also the same as for U_p. The lemmas used for that case work just as well here, but we won't go through the proof.

[1] In fact, a bit more than this is true. If two finite fields have the same size, then they are also isomorphic. This motivates people to write $\mathbb{F}_q$ or $GF(q)$ for "the" field with q many elements in it, but we won't use this notation or this fact. It turns out that there exists a finite field for every $q = p^k$, p prime, so the notation makes sense whenever q is a prime power.

3 Application: Elliptic Curve Cryptography

In this section we will talk about an important application of finite fields to something called elliptic curve cryptography. Before we get there, let's discuss what an elliptic curve is.

3.1 Elliptic Curves

Elliptic curves are a special type of curve. Despite the name, they are *not* ellipses.[2] They can be defined via an equation.

Definition 7. *An elliptic curve E is the set of points in $\mathbb{R}^2$ (the plane) satisfying the equation*

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

for some $A, B \in \mathbb{R}$ with $\Delta := 4A^3 + 27B^2 \neq 0$.

Example. $E : y^2 = x^3 - x$ is an elliptic curve with $A = 0, B = -1$, as $\Delta = 4(0)^3 + 27(-1)^2 \neq 0$.

The condition on Δ is a technical one: it means that when you draw a picture of the set of points of E, it looks smooth (i.e. no sharp corners, etc.). It may not seem important, but for theoretical reasons it makes a big difference.

Elliptic curves have a plethora of very interesting properties that have led mathematicians to study them for centuries. One interesting number theoretic question one can ask about them is: do they have any points with are integers? Or even rational numbers? There is still active research today on finding answers to these questions. We were able to use properties of $\mathbb{Z}[i]$ to resolve the simple case that $E : y^2 = x^3 - 1$ has only one integer solution: (1,0). In general, those techniques don't quite work.

One of the key features elliptic curves have is that you can "add" their points together in a funny way to get another point on the curve. This is different from just adding the respective coordinates of the two points. Instead it comes from a geometric process:

Given $P, Q \in E$, take a line passing through P and Q. Then it will intersect E at another point R. Reflect R over the y-axis. This new point is then the definition of $P + Q$.

At least, that's the general idea. Unfortunately, this won't work for all pairs of points $P, Q \in E$. There are two special cases. The first is if $P = Q$. Then instead of taking a line between P and Q, we should take the tangent line of E at P, and that will intersect E at a unique point R. (Here is actually where we need the condition $\Delta \neq 0$). Then reflect R as before to get $P + P$.

The other special case isn't so easily fixed. This happens if P and Q are vertically above each other, i.e. their x-coordinates are equal. Then the line connecting them is vertical, and won't intersect E at a third point R. To fix this, we actually add another point "at infinity" to E, denoted $\mathcal{O}$. In this special case when P and Q have the same x-coordinate, we define $P + Q = \mathcal{O}$. We extend the addition rule to include $\mathcal{O}$ by $P + \mathcal{O} = P$ for all $P \in E$.

If we use geometry (and a little calculus) to write out equations for the process above, we get the following theorem.

[2]The reason for the weird name is historical: study of these curves was motivated by trying to compute measurements of arclength of ellipses, e.g. like the length of the paths of planetary objects. Centuries later mathematicians have found applications of these curves to many things not involving ellipses in the slightest.

Theorem 8. *Let $P = (x_1, y_1), Q = (x_2, y_2) \in E$. If $x_1 = x_2$ with $y_1 \neq y_2$, then $P + Q = \mathcal{O}$. Otherwise, we have the following formula for $(x_3, y_3) = P + Q$:*

$$x_3 = \lambda^2 - x_1 - x_2$$
$$y_3 = \lambda(x_1 - x_3) - y_1$$

where λ is either

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1} \text{ if } P \neq Q \text{ or } \frac{3x_1^2 + A}{2y_1} \text{ if } P = Q.$$

By looking at the formulas, we see that if A, B defining E actually lie in $\mathbb{Q}$, then whenever P and Q have rational coordinates, so does $P + Q$ (as then so will λ - check this!). This observation is crucial for studying rational points on E.

3.2 Elliptic Curves over Finite Fields

The real fun begins when we try to enact a similar story over a finite field $F = \mathbb{Z}_p[x]_{q(x)}$.

Definition 9. *An elliptic curve E over F is the set of points (x, y) with $x, y \in F$ such that*

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

with $A, B \in F$ and $\Delta := (4A^3 + 27B^2) \neq 0$ in F.[3]

Note that the equality happening above is equality in F. So in practice, it may look like a type of congruence. Sometimes we may write $E(F)$ to describe the points of E over F when the equation written doesn't obviously have coefficients in F (i.e. may be mistaken for integers instead of elements of $\mathbb{Z}_p$).

Example. $E : y^2 = x^3 + 2x + 4$ is an elliptic curve over $F = \mathbb{Z}_5$ since $\Delta = 4(2)^3 + 27(2)^2 \equiv 4 \not\equiv 0 \bmod 5$. We can describe its set of points by just checking which pairs of $x, y \in \mathbb{Z}_5$ fit into the above equation (which is really congruence mod 5). We find that

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

The amazing thing is that the algebraic formulas above still work for adding points even over a finite field, as long as we add an extra point $\mathcal{O}$ to E again playing the same role. This means we might actually write $E(\mathbb{Z}_5) = \{\mathcal{O}, (0, 2), (0, 3), (2, 1), (2, 4), (4, 1), (4, 4)\}$ to include the point at infinity.

Example. With E as above over $\mathbb{Z}_5$, let $P = (0, 2), Q = (4, 1)$. Then $\lambda = \frac{-1}{4} \equiv 1 \bmod 5$. So using the formulas in the theorem, we get $x_3 = 1^2 - 4 - 0 \equiv 2 \bmod 5$ and $y_3 \equiv 1(0 - 2) - 2 \equiv 1 \bmod 5$. So $P + Q = (2, 1)$.

Example. Using the same E, let $P = (4, 1)$. Then to compute $P + P$, we use $\lambda = (3 \cdot 4^2 + 2)/(2 \cdot 1) \equiv 0 \bmod 5$. So $x_3 \equiv 0 - 4 - 4 \equiv 2 \bmod 5$ and $y_3 \equiv -1 \equiv 4 \bmod 5$. Hence $P + P = (2, 4)$.

Example. It's important that we be able to do similar computations not just when $F = \mathbb{Z}_p$, but when F is a finite field like the ones we studied more generally above. For example,

$$E : y^2 = x^3 + x + T$$

[3] A slight remark: This definition is good except when $F = \mathbb{Z}_2[x]_{q(x)}$ or $\mathbb{Z}_3[x]_{q(x)}$, i.e. when $1+1 = 0$ or $1+1+1 = 0$ in F. In these cases, the equations should be modified slightly, but we'll ignore this subtlety for our purposes.

is an elliptic curve over $F = \mathbb{Z}_5[T]_{T^2-2}$ (here we use the variable T instead of x to avoid confusion with the x appearing in the equation for E). We can see that $\Delta = 4 + 27T^2 = 4 + 2 \cdot 2 = 3 \neq 0$ in F, so it fits the definition. We can find points on E by listing all 25 elements of F out and plugging them in to the right side, and seeing if they have square roots.

Similar to when we studied U_p, we can define the *order* of $P \in E(F)$ to be the smallest positive integer n such that $nP := P + P + \cdots + P$ (n-times) $= \mathcal{O}$, the "neutral" element for $E(F)$ relative to this funny addition. This satisfies many properties similar to the order for elements of U_p. For example, the order of any point P always divides $\#E(F)$. (Compare with the order of $k \in U_m$ always divides $\varphi(m) = \#U_m$.)

3.3 Crytography

We've set up enough theory to be able to show a simple application to crytography. Recall the normal Diffie-Hellman key exchange. The goal is for Alice and Bob to obtain a shared secret number (called a key) k even in the presence of an eavesdropper Eve. To do this, we had Alice and Bob decide on a large prime p with a generator g of U_p. Then Alice and Bob pick secret numbers a and b respectively.

Alice sends Bob $A = g^a \bmod p$ and Bob sends Alice $B = g^b \bmod p$. Then each of them compute $A^b = B^a = g^{ab} \bmod p$. This is their shared key k.

The neat thing is that we can take this model and tweak it a little to make a better version of the Diffie-Hellman key exchange, called the Elliptic Curve Diffie-Hellman key exchange, or ECDH.

The ECDH Protocol:
Public: An elliptic curve E over a finite field F; a point P on E with large order.
Private: Alice picks $a > 0$ secretly; Bob picks $b > 0$ secretly.
The Exchange: Alice sends $A = aP = P + P + \cdots + P$($a$-times) to Bob. Bob sends $B = bP = P + P + \cdots + P$ to Alice.
The Key: Alice computes aB and Bob computes bA. Each equal $k = (ab)P$, so this is their shared secret.

How could Eve possibly recover k? Well, of course if she could figure out either a or b, then she can compute k using the public information A or B. The problem of figuring out a given A is called the *Elliptic Curve Discrete Log Problem*, or ECDLP, by analogy with the discrete log problem of determining a given g^a for U_p. The ECDLP is generally believed to be at least as difficult as solving the usual discrete log problem.

For example, when trying to solve the usual discrete log problem for U_p, there is a method known as "index calculus" which is one reasonable attack on the classic DH key exchange if the numbers involved as too small. However, for the ECDLP, there is no known analogue of this attack. Hence it's believed that one can actually use smaller numbers for ECDH to achieve the same degree of security, since one doesn't need to defend against this kind of attack. This makes the ECDH protocol a much more efficient version of the usual Diffie-Hellman key exchange when it comes to computational resources. In fact, the ECDH is used today in practice in a variety of situations, including when connecting securely to some webpages.

One important thing to be aware of with this protocol is that if $\#E(F)$ is a product of small primes, then there are actually efficient ways to solve the ECDLP. That's bad news for Alice and Bob, so they should be careful about choosing their E and F to avoid this.

How many points should E generally have over a finite field? It can only have at most $q^2 + 1$ if F has q elements (including our "point at infinity"), but in fact we can do much better.

Theorem 10. *Let E be an elliptic curve over a finite field with q elements. Then* [4]

$$|\#E(F) - (q+1)| \leq 2\sqrt{q}.$$

Example. If E is an elliptic curve over $\mathbb{Z}_5$, then $(5+1) - 2\sqrt{5} \leq \#E(\mathbb{Z}_5) \leq (5+1) + 2\sqrt{5}$, or $2 \leq \#E(\mathbb{Z}_5) \leq 10$. This is much better than our naive estimate that $1 \leq E(\mathbb{Z}_5) \leq 26$. (!) Note that the lower bound actually says something interesting here: that there's a "genuine" solution to $y^2 \equiv x^3 + Ax + B \bmod 5$ (given $\Delta \neq 0 \bmod 5$) that isn't just $\mathcal{O}$. This too is also very not obvious.

There exist some sophisticated techniques today to be able to compute $\#E(F)$ very quickly for F a finite field, without actually listing out every point of $E(F)$. This is important for ruling out some bad choices for E and F, and $\#E(F)$ is also used in some other algorithms in cryptography.

[4] Here's a neat fact about this amazing theorem, for fans of analytic number theory. There is actually a way to associate a type of zeta function $Z_E(s)$ to an elliptic curve E over F a finite field. It turns out the zeroes of this zeta function satisfy a regularity very similar to that conjectured to hold for the zeroes of $\zeta(s)$, the Riemann zeta function. This analogue of the Riemann hypothesis for $Z_E(s)$ turns out to be equivalent to this theorem, but it's much easier to prove than the classical one which is still unproven! However, this equivalence suggests that the theorem is not in any sense trivial and actually involves a bit of work to prove.