Heidi Lie Williams

7 November 2002

"A Friendly Introduction to Elliptic Curves"

The notes for this lecture have been compiled from the following sources:

- I. Blake, G. Seroussi, and N. Smart, *Elliptic Curves in Cryptography*, Cambridge University Press, 2002.
- J. Silverman, *The Arithmetic of Elliptic Curves*, Springer-Verlag, 1986.
- J. Silverman and J. Tate, *Rational Points on Elliptic Curves*, Springer-Verlag, 1992.
- J. Silverman, *A Friendly Introduction to Number Theory*, Prentice Hall, 1996.
- A. Wiles, "The Birch and Swinnerton-Dyer Conjecture," *Collected Papers.*

My senior thesis this year is exploring the theory of elliptic curves, as well as investigating a computational problem relating to Lenstra's Elliptic Curve Method (ECM) of factorization. My readings this fall have covered the preliminaries of algebraic geometry and the theory of elliptic curves, the second of which I will give an overview of in this lecture. There are two main approaches to the theory of elliptic curves: the first develops many of the basic theorems through complex analytic methods; the second (used here) relies more heavily on techniques from algebraic geometry. Although many of the results in this field rely heavily on techniques in algebraic geometry, one needs no machinery at all to write down the equation of an elliptic curve and to do explicit computations with it.

I. Why are they called elliptic curves?

- First, here are three examples of elliptic curves:

$$E_1 : y^2 = x^3 + 17$$

$$E_2 : y^2 = x^3 + x$$

$$E_3 : y^2 = x^3 - 4x^2 + 16$$

- One of the first things we will notice when studying elliptic curves is that they are **not** ellipses, so it is worth taking a moment here to give a brief account of how this name arises.
- Just as the evaluation of the integral giving arc length on a circle, namely $\int \frac{1}{\sqrt{1-x^2}} dx$, leads to an (inverse) trigonometric function, the analogous problem for arc length of an ellipse leads to an integral of the form $\int \frac{dx}{\sqrt{4x^3 - g_2 x - g_3}}$, which is known as an "elliptic integral."
- In the Weierstrass theory of elliptic functions, it is shown that whenever you have two complex numbers g_2, g_3 such that the polynomial $4x^3 - g_2 x - g_3$ has distinct roots, then you can find complex numbers ω_1, ω_2 (called periods) by evaluating certain definite integrals. These periods are $\mathbb{R}$-linearly independent, and one then looks at the group formed by taking all of their $\mathbb{Z}$-linear combinations:

$$L = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \{n_1\omega_1 + n_2\omega_2 : n_1, n_2 \in \mathbb{Z}\}.$$

L is a subgroup of $\mathbb{C}$ and, in particular, is a lattice.

- One uses these periods to define a function $\wp$ by

$$\wp(u) = \frac{1}{u^2} + \sum_{\substack{w \in L \\ w \neq 0}} \left(\frac{1}{(u-w)^2} - \frac{1}{w^2} \right)$$

where $\wp(u)$ is a meromorphic function that is doubly-periodic; i.e.

$$\begin{aligned} &\forall u \in \mathbb{C} \\ &\wp(u + \omega_1) = \wp(u) \\ &\wp(u + \omega_2) = \wp(u) \\ &\Rightarrow \forall u \in \mathbb{C}, \forall w \in L \\ &\wp(u + w) = \wp(u) \end{aligned}$$

- It turns out that every doubly periodic function with periods which are $\mathbb{R}$-linearly independent satisfies an equation of the form

$$\wp'^2 = 4\wp^3 - g_2\wp - g_3 \qquad (*).$$

For future reference, $\wp$ is referred to as the Weierstrass $\wp$-function.

- If we consider the pair $(\wp, \wp')$ as a point in space, then the solutions to (*) provide a mapping from a torus to the following curve, which is in fact an example of an elliptic curve:

$$y^2 = 4x^3 - g_2 x - g_3 .$$

II. What are elliptic curves?

- In general, elliptic curves are degree three equations of the form

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

where a, b, c are fixed. We are looking for pairs of numbers (x, y) that solve the equation. Note that a more accurate name for an elliptic curve is an "Abelian variety of dimension one."

  - We'll take a moment here to discuss this. The field of algebraic geometry deals with "curves" in any number of dimensions, called algebraic varieties. When these varieties also have a group operation (which we will discuss later), they are called algebraic groups. It turns out there are only two types of algebraic groups:
    1) The first type is a linear algebraic group, which is isomorphic to an algebraic subgroup of the general linear group.
    2) If an algebraic variety has the technical property of being "complete," it is an Abelian variety because the group operation must then be commutative.

    The only algebraic group of both types is the trivial group. Elliptic curves are, as mentioned, Abelian varieties of dimension one.

- We can define elliptic curves more rigorously by using what are known as Weierstrass equations, through which we will be able to use explicit formulas as much as possible to replace the need for general theory.
- A basic theorem says that every elliptic curve has a Weierstrass equation. That is, every elliptic curve can be written as the locus in P^2 of a cubic equation

$$Y^2 Z + a_1 XYZ + a_3 YZ^2 = X^3 + a_2 X^2 Z + a_4 XZ^2 + a_6 Z^3 \text{ where } a_i \in \overline{K} .$$

- To ease notation, we can write this using non-homogeneous coordinates

$$x = \frac{X}{Z}$$
$$y = \frac{Y}{Z}$$

such that

$$E : y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$$

with an extra point $O = [0,1,0]$ at infinity. If $char\left(\overline{K}\right) \neq 2$, then we can let $y = \frac{1}{2}\left(y - a_1x - a_3\right)$ so that we now have

$$E : y^2 = 4x^3 + b_2x^2 + 2b_4x + b_6$$

where we have

$$\begin{aligned} b_2 &= a_1^{\ 2} + 4a_2 \\ b_4 &= 2a_4 + a_1a_3 \,. \\ b_6 &= a_3^{\ 2} + 4a_6 \end{aligned}$$

From here, let us define the following quantities:

$$\begin{aligned} b_8 &= a_1^{\ 2}a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^{\ 2} - a_4^{\ 2} \\ c_4 &= b_2^{\ 2} - 24b_4 \\ c_6 &= -b_2^{\ 3} + 36b_2b_4 - 216b_6 \\ \Delta &= -b_2^{\ 2}b_8 - 8b_4^{\ 3} - 27b_6^{\ 2} + 9b_2b_4b_6 \\ j &= \frac{c_4^{\ 3}}{\Delta} \\ \omega &= \frac{dx}{2y + a_1x + a_3} \end{aligned} \quad .$$

The quantity Δ given above is called the discriminant of the Weierstrass equation, j is called the j-invariant of the elliptic curve E, and ω is the invariant differential associated with the Weierstrass equation.

- Although we will not heavily discuss the *j* – invariant or the invariant differential in this lecture, both are very useful notions. For example, two elliptic curves defined by Weierstrass equations are isomorphic over a field *K* if and only if they have the same *j* – invariant.
- Our definition as to what sort of values x, y represent has been deliberately vague. In the most elementary case, $x, y \in \mathbb{R}$. For arithmetic purposes, we consider $x, y \in \mathbb{Q}$ or $x, y \in \mathbb{F}_p$ for some prime *p*. But the theory of elliptic curves is richest when $x, y \in \mathbb{C}$, not only because of the algebraic closure of $\mathbb{C}$ but also because of the rich analytic theory that exists for complex functions. The equation of an elliptic curve defines *y* as an "algebraic function" of *x*. It turns out that an elliptic curve, defined as a locus of points, is also the Riemann surface associated with the algebraic function defined by the equation. In fact, an elliptic curve is a compact Riemann surface of genus 1; additionally, every compact Riemann surface of genus 1 is an elliptic curve.

- Topologically, an elliptic curve is equivalent to a torus. This topological equivalence is given by an explicit mapping involving the Weierstrass $\wp$-function and its first derivative: namely, $z \mapsto \left(\wp(z), \wp'(z)\right)$. We also note that if E is an elliptic curve over $\mathbb{C}$, then E is a complex Lie group (i.e. a complex manifold with a group law given locally by complex analytic functions).

III. Studying solutions to elliptic curves.

- One way to find rational solutions to a given elliptic curve is, of course, by trial and error: simply plug in small values of *x* and see if $x^3 + ax^2 + bx + c$ is a perfect square. However, we can use a geometric method to find or create rational solutions more efficiently.
- **Example.** For $E_1 : y^2 = x^3 + 17$, let us draw lines though $P = \left(-2, 3\right)$ and see what other points we can find. Try the line with slope 1, i.e. $y - 3 = x + 2$. To find the intersection of this line with E_1, substitute $y = x + 5$ into E_1 as follows:

$$\begin{aligned} &y^2 = x^3 + 17 \\ &\left(x+5\right)^2 = x^3 + 17 \\ &0 = x^3 - x^2 - 10x - 8 \end{aligned}$$

We know we could find the roots of this cubic, but it is easier to note that we already know one of the solutions since E_1 and our line both go through $P = \left(-2, 3\right)$. Therefore, $x = -2$ must be a root, which implies

$$0 = \left(x+2\right)\left(x^2 - 3x - 4\right)$$

and we easily see that this quadratic polynomial has roots at $x = -1, 4$. Substituting these values into $y = x + 5$ gives two new points on E_1, namely $\left(-1, 4\right), \left(4, 9\right)$.

- **Another example.** Easy enough? Let's consider one more example. For $E_1 : y^2 = x^3 + 17$, take the line through $P = \left(-2, 3\right)$ with slope 3, i.e. $y - 3 = 3(x + 2)$. As before, we use substitution, this time with $y = 3x + 9$, which gives us:

$$\begin{aligned} &y^2 = x^3 + 17 \\ &\left(3x+9\right)^2 = x^3 + 17 \\ &0 = x^3 - 9x^2 - 54x - 64 \\ &0 = \left(x+2\right)\left(x^2 - 11x - 32\right) \end{aligned}$$

This quadratic has roots $x=\frac{11\pm\sqrt{249}}{2}$. This is, of course, not what we were hoping for because we want to have $x,y\in\mathbb{Q}$. Where was the problem? Suppose we draw a line *L* of slope *m* passing through $P=(-2,3)$ and find the intersection of *L* with E_1. Substituting $y=m(x+2)+3$ into E_1 gives the following:

$$\begin{aligned}
&y^2=x^3+17\\
&\left(m(x+2)+3\right)^2=x^3+17\\
&0=x^3-m^2x^2-(4m^2+6m)x-(4m^2+12m-8)\\
&0=(x+2)\left(x^2-\left(m^2+2\right)x-\left(2m^2+6m-4\right)\right)
\end{aligned}$$

Note that this quadratic polynomial is unlikely to have rational roots. How can we compel this polynomial to have rational roots? If we force our original cubic to have two rational roots, by the Rational Roots Theorem the third will have to be rational also. Therefore, we should choose a line that already intersects E_1 at two rational points.

- **One more example.** Therefore, start with $P=(-2,3)$ and $Q=(2,5)$ on E_1. The line connecting *P*, *Q* has slope $\frac{5-3}{2+2}=\frac{1}{2}$. Therefore $y=\frac{x}{2}+4$ and thus we have:

$$\begin{aligned}
&y^2=x^3+17\\
&\left(\frac{1}{2}x+y\right)^2=x^3+17\\
&0=x^3-\frac{1}{4}x^2-4x+1\\
&0=(x-2)(x+2)\left(x-\frac{1}{4}\right)
\end{aligned}$$

which implies that $y=\frac{33}{8}$. Therefore $R=\left(\frac{1}{4},\frac{33}{8}\right)$ is a rational solution. We can then repeat this process using *R* with *P* or *Q*. Continuing in this fashion, it turns out that there exist infinitely many rational points on E_1.

- Problems on curves of genus 1 are prominently featured in Diophantus' *Arithmetica*, in which he implicitly uses our geometric method but does not iterate it. Fermat was the first to realize that this geometric method could sometimes be used to find infinitely many solutions.

- However, a main theorem in the theory of elliptic curves tells us that even the infinitely many solution to E_1 can be created from a finite generating set.
- **Mordell's Theorem.** Let *E* be an elliptic curve given by $y^2 = x^3 + ax^2 + bx + c$ where $a, b, c \in \mathbb{Z}$ such that $\Delta(E) \neq 0$. Then there exists a finite list of solutions $P_1, \ldots, P_r$ with rational coordinates such that every rational solution can be obtained from this finite list through our geometric method.
- Mordell conjectured more generally that the set of rational points on any algebraic curve of genus greater than 1 is finite; this conjecture was proven by the German mathematician Gerd Faltings of Princeton, who received the Fields Medal primarily for this proof of Mordell's conjecture.
- In 1922, Mordell noted, "Mathematicians have been familiar with very few questions for so long a period with so little accomplished in the way of general results, as that of finding the rational points on elliptic curves."
- Andrew Wiles of Princeton has noted, "As of yet, the proof [of Mordell's Theorem] is not effective," meaning that the proof does not produce an algorithm with which to find the rational points. There exists an effective bound, but this is not sufficient.
- It turns out that this geometric operation on the points of an elliptic curve transforms the curve into an Abelian group. We can now define this group operation more rigorously as follows.
- **Composition Law.** For $O = [0,1,0]$, let $P, Q \in E$, *L* the line connecting *P* and *Q*, and *R* the third point of intersection of *L* with *E*. Let L' denote the line connecting R and *O*. Then $P \oplus Q$ is the point such that L' intersects *E* at *R*, *O*, and $P \oplus Q$.
- **Proposition.** The composition law has the following properties:
  - If a line *L* intersects *E* at the (not necessarily distinct) points *P*, *Q*, and *R*, then $(P \oplus Q) \oplus R = 0$.
  - $\forall P \in E, P \oplus 0 = P$.
  - $\forall P, Q \in E, P \oplus Q = Q \oplus P$.
  - Let $P \in E$. Then $\exists (-P) \in E$ such that $P \oplus (-P) = 0$.
  - Let $P, Q, R \in E$. Then $(P \oplus Q) \oplus R = P \oplus (Q \oplus R)$.

  Thus the composition law makes *E* into an Abelian group with the identity element $O = [0,1,0]$.
- Back to our problem of counting points, let us consider our example E_2.

- **<u>Proposition.</u>** The elliptic curve $E_2 : y^2 = x^3 + x$ has only one rational solution, namely $(0,0)$.

  Sketch of proof. Let $\left(\frac{A}{B}, \frac{C}{D}\right)$ be a rational point on E_2 in lowest terms with $B, D > 0$. We can show $A = 0 = C$ simply be using divisibility rules derived from the fact that

$$\begin{aligned}&\left(\frac{C}{D}\right)^2 = \left(\frac{A}{B}\right)^3 + \left(\frac{A}{B}\right)\\ &\Rightarrow\\ &C^2B^3 = A^3D^2 + AB^2D^2\end{aligned}$$

  There is no need to use this time to go over the simple algebraic manipulations of this proof, but that is the main idea.

- Now consider $E_3 : y^2 = x^3 - 4x^2 + 16$. Trial and error can find four points on E_3: $P_1 = (0,4), P_2 = (4,4), P_3 = (0,-4), P_4 = (4,-4)$. Using our geometric method, the line connecting P_1, P_2 has equation $y = 4$. To find the intersection with E_3,

$$\begin{aligned}&4^2 = x^3 - 4x^2 + 16\\ &0 = x^3 - 4x^2 = x^2(x-4)\end{aligned}$$

  This implies that *x*=0 is a double root. Therefore the line connecting P_1, P_2 has no other points of intersection with E_3. Similar occurrences arise when we choose any two of $P_1 = (0,4), P_2 = (4,4), P_3 = (0,-4), P_4 = (4,-4)$. It turns out that these are the only four rational points we can find through our geometric method, which thus motivates the following definition.

- **<u>Defn.</u>** A finite collection of points $P_1, \ldots, P_r$ ($r \geq 3$) on an elliptic curve $y^2 = x^3 + ax^2 + bx + c$ is called a <u>torsion collection</u> if the set cannot be enlarged by using our geometric method.

- **<u>Theorem.</u>** (proven independently by Nagell, Lutz). Let *E* be an elliptic curve given by $y^2 = x^3 + ax^2 + bx + c$ where $a, b, c \in \mathbb{Z}$, $\Delta(E) \neq 0$ and let $P_1, \ldots, P_r$ be a torsion collection on *E* consisting of rational points. Then for a non-zero torsion point $P_i = (x_i, y_i)$, we have that $x_i, y_i \in \mathbb{Z}$ and $y_i^2 | 16\Delta(E)$.

- Note that this theorem implies that all rational points in a torsion collection have integer coordinates.

- A recently proven (and difficult!) theorem by Barry Mazur of Harvard shows that a torsion collection contains at most 15 points, and in fact must be one of 2, 3, 4, 5, 6, 7, 8, 9, 11, or 15. Additionally, Mazur proved that the order of an element in a torsion collection must be ≤ 12 (and not =11).
- The Nagell-Lutz Theorem can lead us into another result about solutions in the integers.
- **Theorem.** (Siegal) Let *E* be an elliptic curve given by $y^2 = x^3 + ax^2 + bx + c$ where $a, b, c \in \mathbb{Z}$ and $\Delta(E) \neq 0$. Then there exist only finitely many solutions such that $x, y \in \mathbb{Z}$.

IV. Studying solutions to elliptic curves over finite fields.

- Because it can be difficult to solve a Diophantine equation, we can choose to treat the equation as a congruence and look for solution (mod *p*).
- The most important arithmetic quantity associated with an elliptic curve defined over a finite field is its number of rational points.
- **Example.** Let us find all the solutions (mod 7) to $x^2 + y^2 = 1$, i.e. we wish to solve the congruence $x^2 + y^2 \equiv 1 (\text{mod}\, 7)$. We can simply try pairs (*x*,*y*) such that $0 \leq x, y \leq 6$ and see which work: this yields the full set of solutions, which is $(0,1),(0,6),(1,0),(2,2),(2,5),(5,2),(5,5),(6,0)$. Therefore there exist eight solutions (mod 7). Similarly, there exist twelve solutions (mod 11).
- This type of pattern can lead us to look at elliptic curves and count how many points they have (mod *p*) for various primes *p*. Define N_p to be the number of points (mod *p*).
- **Example.** Consider $E_2 : y^2 = x^3 + x$.

p	N_p
2	2
3	3
5	3
7	7
11	11
13	19
17	15
19	19

- The number of points modulo *p* on an elliptic curve exhibits many wonderful and subtle patterns. For many *p*, $N_p = p$. In fact, aside from *p*=2, the primes such that $N_p = p$ are exactly the set of primes less than 71 which are congruent to 3 (mod 4). Although we will not give a proof of this fact, we can take a moment here to consider why it is true. In general, if we are trying to find the solutions (mod *p*) to $y^2 = x^3 + ax^2 + bx + c$, we substitute $x = 0,1,\ldots, p-1$ and check, for each *x*, whether $x^3 + ax^2 + bx + c$ is a perfect square. We would expect these values to be squares about half the time, since half the numbers $x = 0,1,\ldots, p-1$ are quadratic residues. Also observe that if $x^3 + ax^2 + bx + c \equiv t^2 (\text{mod } p)$, for some *p*, then $y = \pm t$. Therefore approximately half the values of *x* give two solutions (mod *p*) and half of the values give no solution (mod *p*). Therefore there exist $2 \times \frac{1}{2} p = p$ solutions.
- From here, we can define a quantity which will allow us to continue to investigate the difference between *p* and N_p.
- **<u>Defn.</u>** Define $a_p = p - N_p$ to be the <u>*p*-defect</u>, also known as the "trace of Frobenius."
- Some elliptic curves have patterns which are currently being uncovered and discovered.
- Noam Elkies of Harvard proved (1987) that for every elliptic curve, there exist infinitely many primes such that $a_p = 0$.
- In general it is a very difficult problem to find patterns in the a_p's, but in special cases we can discuss some results. For example, an elliptic curve is said to have a <u>complex multiplication</u> if it satisfies a certain sort of transformation property. It is known that elliptic curves with complex multiplication have half of their a_p's equal to 0.
- The following theorem was conjectured by E. Artin in his thesis and proved by Hasse in the 1930's.
- **<u>Hasse's Theorem.</u>** Let N_p be the number of points (mod *p*) on *E*, and $a_p = p - N_p$ the *p*-defect. Then

$$|a_p| < 2\sqrt{p}.$$