Lecture 25: Elliptic Curves 1: Introduction

William Stein

Math 124 HARVARD UNIVERSITY **Fall 2001**

1 The Definition

Finally we come to elliptic curves, which I think are the most exciting and central *easily accessibly* objects in modern number theory. There are so many exciting things to tell you about elliptic curves, that the course is suddenly going to move more quickly than before.

Definition 1.1. An *elliptic curve* E over a field K is a plane cubic curve of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

where $a_1, a_2, a_3, a_4, a_6 \in K$ and

$$\Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6 \neq 0,$$

where

$$b_2 = a_1^2 + 4a_2, \quad b_4 = 2a_4 + a_1a_3, \quad b_6 = a_3^2 + 4a_6.$$

Help! Don't worry, when 2 and 3 are not equal to 0 in K, using completing the square and a little algebra we find a change of coordinates that transforms the above cubic equation into the form

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

and then $\Delta = -16(4a^3 + 27b^2)$. We will consider only elliptic curves of the form $y^2 = x^3 + ax + b$ for a while.

Hey! That's not an ellipse! You're right, elliptic curves are *not ellipses*; they are curves that first arose when 19th century mathematicians studied integral formulas for the arc lengths of ellipses.

In these lectures, I'll give you a glimpse into two main ways in which elliptic curves feature in mathematics. On the left hand, they provide the simplest example of a class of diophantine equations that we still can't totally solve. On the right hand, when K is a finite field (or, more sneakily, a finite ring), elliptic curves can be used as a tool for both making and breaking cryptosystems.

2 Linear and Quadratic Diophantine Equations

Consider the following question:

> Let $F(x, y)$ be an irreducible polynomial in two variables over $\mathbb{Q}$. Find all rational numbers x_0, y_0 such that $F(x_0, y_0) = 0$.

When F is linear, this problem is easy. The equation

$$F(x, y) = ax + by + c = 0$$

defines a line, and letting $y = t$, the solutions are

$$\left\{ \left(-\frac{b}{a}t - \frac{c}{a}, t \right) \; : \; t \in \mathbb{Q} \right\}.$$

When F is quadratic, the solution is not completely trivial, but it is well understood. In this case, the equation $F = 0$ has infinitely many rational solutions if and only if it has at least one solution. Moreover, it is easy to describe all solutions when there is one. If (x_0, y_0) is a solution and L is a non-tangent line through (x_0, y_0), then L will intersect the curve $F = 0$ in exactly one other point (x_1, x_1). Also $x_1, y_1 \in \mathbb{Q}$ since a quadratic polynomial over $\mathbb{Q}$ with 1 rational root has both roots rational. Thus the rational points on $F = 0$ are in bijection with the slopes of lines through (x_0, y_0).

Chapter 2 of [Kato et al.] is about how to decide whether or not an F of degree 2 has a rational point. The answer is that $F = 0$ has a rational solution if and only if $F = 0$ has a solution with $x_0, y_0 \in \mathbb{R}$ and a solution with $x_0, y_0 \in \mathbb{Q}_p$ for every "p-adic field" $\mathbb{Q}_p$. This condition, though it might sound foreboding, is easy to check in practice. I encourage you to flip through chapter 2 of loc. cit.

3 Points on Elliptic Curves

Next suppose that F is an irreducible cubic polynomial. The question of whether or not $F = 0$ has a rational solution is still an *open problem*! We will not consider this problem further until we discuss the Birch and Swinnerton-Dyer conjecture.

Suppose that $F = 0$ has a given rational solution. Then one can change coordinates so that the question of finding the rational solutions to $F = 0$ is equivalent to the problem of finding all rational points on the elliptic curve

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

Recall that when F has degree 2 we can use a given rational point P on the graph of $F = 0$ to find all other rational points by intersecting a line through P with the graph of $F = 0$. The graph of $y^2 = x^3 + ax + b$ looks like

[egg and curvy line] or [curvier line]

Notice that if P is a point on the graph of the curve, then a line through P (usually) intersects the graph in exactly *two* other points. In general, these two other points usually do not have rational coordinates. However, if P and Q are rational points on the graph of $y^2 = x^3 + ax + b$ and L is the line through P and Q, then the third point of intersection with the graph will have rational coordinates. Explicitly, if $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ then the third point of intersection has coordinates[1]

$$x_3 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2 - x_1 - x_2, \qquad y_3 = \frac{y_2 - y_1}{x_2 - x_1}x_3 - \frac{y_2x_1 - y_1x_2}{x_2 - x_1}.$$

Thus, given two points on E, we can find another. Also, given a single point, we can draw the tangent line to E through that point and obtain a third point.

3.1 To Infinity!

At first glance, the above construction doesn't work if $x_1 = x_2$. [draw picture]. Fortunately, there is a natural sense in which the graph of E is missing one point, and when $x_1 = x_2$ this one missing point is the third point of intersection.

The graph of E that we drew above is a graph in the plan $\mathbb{R}^2$. The plane is a subset of the projective plane $\mathbb{P}^2$, which I will define in just a moment. The closure of the graph of $y^2 = x^3 + ax + b$ in $\mathbb{P}^2$ has exactly one extra point, which has rational coordinates, and which we denote by ∞. Formally, $\mathbb{P}^2$ can be viewed as the set of triples (a, b, c) with a, b, c not all 0 modulo the equivalence relation

$$(a, b, c) \sim (\lambda a, \lambda b, \lambda c)$$

for any nonzero λ. Denote by $(a : b : c)$ the equivalence class of (a, b, c). The closure of the graph of $y^2 = x^3 + ax + b$ is the graph of $y^2z = x^3 + axz^2 + bz^3$ and the extra point ∞ is $(0 : 1 : 0)$.

Venerable Problem: Find an algorithm that, given an elliptic curve E over $\mathbb{Q}$, outputs a complete description of the set of rational points (x_0, y_0) on E.

This problem is difficult. In fact, so far it has stumped everyone! There is a conjectural algorithm, but nobody has succeeded in proving that it is really an algorithm, in the sense that it terminates for any input curve E. Several of your profs at Harvard, including Barry Mazur, myself, and Christophe Cornut (who will teach Math 129 next semester) have spent, or will probably spend, a huge chunk of their life thinking about this problem. (Am I being overly pessimistic?)

How could one possible "describe" the set of rational points on E in the first place? In 1923, Louis Mordell proved an amazing theorem, which implies that there is a reasonable way to describe the rational points on E. To state his theorem, we introduce the "group law" on E.

[1] It is traditional in a course like ours for me to derive these formulas. I'm not going to, because it's simple algebra and once you see the geometric picture it is easy to carry out. You should do this as an exercise, or read the derivation in [Kato et al.] or [Davenport].

4 The Group Law

Consider the set $E(\mathbb{Q}) = \{\infty\} \cup \{(x_0, y_0) : y_0^2 = x_0^3 + ax_0 + b\}$. There is a natural way to endow the set $E(\mathbb{Q})$ with a *group* structure. Here's how it works. First, the element $\infty \in E(\mathbb{Q})$ is the 0 element of the group. Next, suppose P and Q are elements of $E(\mathbb{Q})$. Just like we did earlier, draw the line through P and Q and let $R = (x_3, y_3)$ be the third point of intersection. Define $P + Q = (x_3, -y_3)$. There are various special cases to consider, such as when $P = Q$ or the third point of intersection is ∞, but I will let your read about them in [Kato et al.]. It is clear that this binary operation on $E(\mathbb{Q})$ satisfies $P + Q = Q + P$. Also, the inverse of $P = (x_1, y_1)$ is $-P = (x_1, -y_1)$. The only other axiom to check in order to verify that $+$ gives $E(\mathbb{Q})$ an abelian group structure is the associative law. This is simple but *very tedious* to check using only elementary methods[2]. Fortunately, we can coerce the computer algebra system MAGMA into verifying the associative law for us:

```
// The field K = Q(a,b,x0,x1,x2)
K<a,b,x0,x1,x2> := FieldOfFractions(PolynomialRing(Rationals(),5));
// The polynomial ring R = K[y0,y1,y2]
R<y0,y1,y2> := PolynomialRing(K,3);
// A maximal ideal of R.
I := ideal<R | y0^2 - (x0^3+a*x0+b), y1^2 - (x1^3+a*x1+b), y2^2-(x2^3+a*x2+b)>;
// The field L contains three distinct "generic" points on E.
L  := quo<R|I>;
E  := EllipticCurve([L| a,b]);  // The elliptic curve y^2 = x^3 + a*x + b.
P0 := E![L|x0,y0]; P1 := E![L|x1,y1]; P2 := E![L|x2,y2];
lhs := (P0 + P1) + P2;  rhs := P0 + (P1 + P2);
lhs eq rhs;
true     // yeah!
```

5 Mordell's Theorem

Theorem 5.1 (Mordell). *The group $E(\mathbb{Q})$ is finitely generated.*

This means that there are points $P_1, \ldots, P_r \in E(\mathbb{Q})$ such that every element of $E(\mathbb{Q})$ is of the form $n_1P_1 + \cdots + n_rP_r$ for some $n_1, \ldots n_r \in \mathbb{Z}$. I won't prove Mordell's theorem in this course. You can find an elementary proof of most of it in §1.3 of [Kato et al.].[3]

Example 5.2. Consider the elliptic curve E given by $y^2 = x^3 + x + 1$. Then $E(\mathbb{Q}) \approx \mathbb{Z}$ with generator $(0, 1)$. We have $2(0, 1) = (-1/4, -9/8)$, $3(0, 1) = (72, 611)$, and $4(0, 1) = \left(-\frac{287}{1296}, \frac{40879}{46656}\right)$.

[2]The right way to prove that the associate law holds is to develop the theory of algebraic curves and define the group law in terms of divisors; this is way outside the scope of this course.

[3]Matt Baker is teaching a graduate course (255r) this semester, and he is just about to present a proof of Weil's generalization of Mordell's theorem.