Elliptic Curves

David Wright Escott

May 24, 2004

Final Paper for Mathematics 129:

Algebraic Number Theory
Taught by Professor William Stein

1 Elliptic Curves

Elliptic curves are found at the intersection of a broad range of mathematics. As such there are definitions of elliptic curves springing from abstract number theory where they are defined in terms of sheafs and schemes, and therefore almost completely dissociated from the underlying field, to complex analysis where elliptic curves found their name as being associated to elliptic integrals. However one's first introduction to an elliptic curve often fails to explain why such a broad range of theory exists for such a simple object. Most commonly one says that an elliptic curve is given by the solutions to a nonsingular cubic in two variables x and y over a field. Usually this field is $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{Q}(\alpha)$ or $\mathbb{Q}_p$. These being the most common fields of use in number theory and algebraic geometry. Frequently one is shown pictures of common elliptic curves such as $y^2 = x^3 - 4x$, $y^2 = x^3 - 3x + 3$, and $y^2 = x^3 - x$:

$y^2 = x^3 - 4x$ $\qquad$ $y^2 = x^3 - 3x + 3$ $\qquad$ $y^2 = x^3 - x$

The reason for this broad range of definitions is that elliptic curves while defined as algebraic curves in projective space, have a natural group structure which encourages the development of the algebraic approaches to the subject. This paper will introduce the group law and will also present a rough outline of a theorem by Hasse which shows that every elliptic curve over $\mathbb{F}_p$ has a rational point. The Proof of this theorem requires a great deal of machinery from a broad range of mathematics. In particular it requires elements from Algebraic Geometry, Differential Geometry, Galois Theory of finite fields, the group structure of elliptic curves, and the group structure of maps between elliptic curves.

2 Weierstrass Equations

Since elliptic curves are cubics in x and y the general form is rather complicated.

$$y^2 + axy + by = x^3 + cx^2 + dx + e$$

is the general form for such a curve, where one actually means to take this as a curve in projective space:

$$Y^2Z + aXYZ + bYZ^2 = X^3 + cX^2Z + dXZ^2 + eZ^3$$

Then one can complete the square $y \to \frac{1}{2}(y - ax - c)$ to get

$$(\frac{1}{2}(y-ax-b))^2 + ax\frac{1}{2}(y-ax-b) + b\frac{1}{2}(y-ax-b) = x^3 + cx^2 + dx + e$$
$$y^2 - a^2x^2 - 2abx - b^2 = \dots$$
$$y^2 = 4x^3 + (4c + a^2)x^2 + (4d + 2ab)x + 4e + b^2$$

which gives the more common form $y^2 = ax^3 + bx^2 + cx + d$. For large x the curve goes as $y^2 = ax^3$ so $y = \pm ax^{3/2}$. The positive part is seen to go straight up and the negative part straight down as x approaches infinity. In projective space we see that the curve is approaching the point the point $(X, Y, Z) = (0, 1, 0)$, since the line $(0, t, 0)$ is vertical in the $x - y$ sense and satisfies $Y^2Z = aX^3 + bX^2Z + CXZ^2 + dZ^3$ since both the left and right sides are zero. This point is called the "point at infinity" or the "distinguished point." At any point in our affine space we can find the point $(0, 1, 0)$ by taking the vertical line through that point and following it to "infinity." This distinguished point is critical in the definition of the group law, and is often denoted O.

3 The Group Law

Given an elliptic curve $E : y^2 = ax^3 + bx^2 + cx + d$ then one can define a group law on points in the curve. If $P, Q \in E$ then $P + Q$ is given by first taking the line from P to Q. Since the curve E is nonsingular and cubic, any line intersecting it has three points of intersection when counted with multiplicity. This third point R is "opposite" from the point $P + Q$. The point $P + Q$ is the third point of intersection of the line RO with E.

To see that this is a group law we note that the distinguished point O is the identity. Letting R be the third point of intersection of E with PO then clearly PO and RO will coincide, but then clearly P is the third point on the line RO so $P + O = P$. Furthermore $+$ is clearly commutative since the lines PQ and QP coincide. The additive inverse $-P$ is given by the third point of intersection of E with PO. By construction the line $P(-P)$ has as its third point O, and then the line OO lying completely within the plane of $Z = 0$ has as its third point of intersection the point O, so $P + -P = O$ is the additive inverse.

In general associativity is hard to prove and requires either the Riemman-Roch theorem or a very unenlightening computation. I will work out an example for the curve $y^2 = x^3 + 17$. This curve has a number of integral points which will simplify the computation. Let $P_1 = (-2, 3)$ $P_2 = (-1, 4)$ $P_3 = (2, 5)$ $P_4 = (4, 9)$ $P_5 = (8, 23)$. Then we can compute $P_1 + P_2 + P_5$. P_1P_2 is given by $(t-2, t+3)$ the roots are then given by $t^2 + 6t + 9 = t^3 - 6t^2 + 12t - 8 + 17$ or $0 = t^3 - 7t^2 + 6t = t(t^2 - 7t + 6) = t(t-1)(t-6)$. The roots $t = 0, t = 1$ are the points P_1, P_2. The point $t = 6$ is $(4, 9) = P_4$. We then take the opposite point $-P_4 = (4, -9) = P_1 + P_2$. The line through $-P_4$ and P_5 is given by $(t+4, 8t-9)$ and the roots are given by $64t^2 - 144t + 81 = t^3 + 12t^2 + 48t + 64 + 17$ or $0 = t^3 - 52t^2 + 192t = t(t^2 - 52t + 192) = t(t-4)(t-48)$, where the third point is $t = 48$ or $(52, 375)$. So the final sum $(P_1 + P_2) + P_5 = (52, -375)$.

Performing the reverse sum $P_1 + (P_2 + P_5)$ we have the line $P_2P_5 = (9t - 1, 19t + 4)$. The intersections are given by the roots of $361t^2 + 152t + 16 = 729t^3 - 243t^2 + 27t - 1 + 17$ or $0 = 729t^3 - 604t^2 - 125t = t(t-1)(729t + 125)$ which gives the third point $(-\frac{206}{81}, \frac{541}{729})$. Therefore $P_2 + P_5 = Q = (-\frac{206}{81}, -\frac{541}{729})$. The line QP_1 is given by $(\frac{-44}{81}t - 2, \frac{-2728}{729}t + 3)$. The points of intersection come from the roots.

$$\frac{7441984}{531441}t^2 - \frac{5456}{243}t + 9 = -\frac{85184}{531441}t^3 - \frac{3872}{2187}t^2 - \frac{176}{27} - 8 + 17$$
$$0 = t(t-1)\left(\frac{-85184}{531441}t - \frac{3872}{243}\right)$$

So $t = -\frac{2187}{22}$ plugging this in to our formula for QP_1 we have $(52, 375)$ which gives the sum $P_1 + (P_2 + P_5) = (52, -375) = (P_1 + P_2) + P_5$.

4 Rational Points

The existence of rational points on elliptic curves is of particular interest. As we have already seen rational points remain rational under the group law. This is a simple corollary of the method used to solve for the remaining root. If we are given $P = (P_x, P_y)$ and $Q = (Q_x, Q_y)$ rational points then the line $PQ = ((P_x - Q_x)t + Q_x, (P_y - Q_y)t + Q_y)$ and therefore PQ_x and PQ_y are polynomials in t with coefficients in $\mathbb{Q}$. Plugging these into E where E has rational coefficients we have a relationship $f(t) = g(t)$ where $f, g \in \mathbb{Q}[t]$. But then since we know two of the roots namely $t = 0, 1$ we can easily solve for the remaining root which must have t rational and therefore must be a rational root.

The presence of this group law and and the closure of rational roots under addition allows the easy computation of new rational roots from others, which greatly aides the search for integral solutions to elliptic curves. In particular it can be shown that the group of rational points is finitely generated and that the set of integral points is finite. These theorems are much too difficult to prove here. Even Hasse's Theorem showing that elliptic curves over finite fields have rational roots requires a great deal of machinery.

5 Hasse's Theorem

Let $K = \mathbb{F}_q$ and let $E(K)$ given by $y^2 = x^3 + ax^2 + bx + c$ be an elliptic curve over K. It is easy to establish an upper bound on the number of rational points in E. There are q choices for x and since $y^2 = x^3 + ax^2 + bx + c$ there are at most 2 choices for y for a given x. This gives a total of $2q$ points in the affine plane given by $Z = 1$. Together with the point at infinity there are at most $2q + 1$ points in $E(K)$. In general however we will only have a point when $f(x) = x^3 + ax^2 + bx + c$ is a perfect square in K. Since $n^2 = -n^2$ only half the numbers in K are perfect squares, and therefore we should expect half of the values of x to yield points on E. So on average there will be approximately q points in $E(K)$.

Theorem 5.1 *Hasse's Theorem. The number of rational points in $E(K)$ is bounded by $q + 1 \pm 2\sqrt{q}$.*

Before we can approach this theorem a great deal must be introduced.

6 Algebraic Geometry

The proof of this theorem uses a good deal of Algebraic Geometry. In particular the use of function fields and local rings on projective varieties is critical to many of the underlying theorems about maps between varieties. In essence the structure of a variety is encapsulated in the field of rational functions defined on that variety, and the local structure at a point in the ring of non-vanishing polynomials defined on the variety near that point. Formally these are constructed as follows.

$K[X]$ is the ring of rational functions on the variety X. This may be affine space or in this case projective space. A variety of zeros (or variety) $V \subset X$ is simply the mutual zeros of a set of functions $\{f_i\}$. Clearly if p is a zero of f_i and f_j then p is a zero of $f_i + f_j$ similarly if p is a zero of f_i then p is a zero of any multiple $f_i g$ of f_i by another polynomial $g \in K[X]$. So we can take $\{f_i\}$ to be the ideal generated by the f_i. Hilbert's Nullstellenstaz establishes an exact correspondence between varieties and ideals over Algebraically closed fields.

Since $I(V)$ the ideal of functions vanishing on V is an ideal in $K[X]$ we can take $K[X]/I(V)$ which is the ring of coordinate functions on V $K[V]$. For any point $p \in V$ there is an ideal in $K[V]$ of functions vanishing on p. Localizing about this ideal gives $K[V]_p$ the local ring at p. Similarly the field of rational of fractions of $K[V]$ is the field of rational functions of V denoted $K(V)$.

This use of Algebraic Geometric constructions allows a great deal to be proven by considering maps between varieties as maps between the associated fields of rational functions. In particular we can define the degree of a map via the pull back. Given $\phi : C_1 \to C_2$ a map between curves (a projective variety of dimension one) there is the associated pull back map $\phi^* K(C_2) \to K(C_1)$ where $\phi^*(f) = f \circ \phi$. The degree of the map ϕ is simply $[K(C_1) : \phi^* K(C_2)]$ as fields spaces, constant maps having degree zero. From this we can define separability and inseparability for maps, by way of their corresponding field extensions.

In particular since elliptic curves are Abelian groups it is useful to consider those rational maps (respecting Algebrao-Geometric properties) between elliptic curves that also preserve the groups themselves. Strong statements can be made about the Galois theory of the function fields of these maps, know as isogenies. In particular for separable maps ϕ the cardinality of the kernel of ϕ is the degree of ϕ.

Theorem 6.1 *If ϕ is separable then $\# \ker \phi = \deg \phi$.*

The proof relies on a result in algebraic geometry. Namely that a rational map has only finitely many points where the map has "poles" in the sense that the the number of points in the preimage of a point p is not the separable degree of ϕ. What this means is that if $K(C_1) : \phi^* K(C_2)$ is separable of degree d then over a function $f \in K(C_2)$ we see exactly d different rational functions $f_i \in K(C_1)$. As a result the zeros of f correspond to the d different zeros of the functions f_d, and so the varieties are a d to 1 covering at most points. At finitely many points there may be kinks which prevent this map from having full degree. However if the map respects the group structure, then all inverse images must have the same cardinality since picking a particular element R in E_1 such that $\phi(R) = Q - P$ we have $\phi^{-1}(P) + \phi^{-1}(Q - P) = \phi^{-1}(Q)$ or $\phi^{-1}(P) + R = \phi^{-1}(Q)$. If $\#\phi^{-1}(P) > \#\phi^{-1}(Q)$ then we would have a violation of the group law in E_1 since n-elements translated by R would become m-elements where $n > m$ so for some $A, B \in \phi^{-1}(P)$ we would have $A + R = B + R$ but not $A = B$.

7 Frobenius Map

There is a particularly useful map which is separable and respects the group structure. This map is known as the Frobenius map. Given an projective variety V over K a field of characteristic p where V is given by a set of homogeneous polynomials $\{f\}$ we can define the qth Frobenius map by $\phi_q : V \to V^q$ by $(X, Y, Z) \mapsto (X^q, Y^q, Z^q)$ where $q = p^r$. This map is called the Frobenius map, and the resulting variety is contained in the variety of roots of $\{f^q\}$ the polynomials found by raising all the the coefficients of f to their qth powers. To see this simply note that $(f(X, Y, Z))^q = a_1^q(X^q)^{m_1} + \cdots + a_n^q(Z^q)^{m_n} = f^q(X^q, Y^q, Z^q)$ since the cross terms will all have coefficients divisible by q and therefore be zero. But then clearly $Z(f^q) \supset V^q$. Now if the field K is the finite field $\mathbb{F}_q$ then the qth power map is the identity on K, so $V^q = V$ and $f^q = f$.

Since the Frobenius map is the identity when restricted to K it is clear that if $P \in E(K)$ then $\phi(P) = P$. Therefore $(\phi - 1)(E(K)) = 0$ so the elements in $E(K)$ are roots of the map $\phi - 1$. We for reasons that will be clear in a moment we would like the reverse statement to hold as well, that is if $\phi(P) = P$ then $P \in E(K)$. This is indeed true for any algebraic extension L of K. A direct corollary of the classification of finite fields is that the finite extensions of K are all extensions of the form $x^{q^r} - 1$ and therefore have Galois Group generated by the Frobenius map. As a result the Frobenius map acts non-trivially on $L \setminus K$ by sending elements to their Galois conjugates and acts as the identity on K itself. Therefore the kernel of $\phi - 1 : L \to L$ is K since if $\phi(P) = P$ then $(\phi - 1)(P) = 0$.

Finally with all those results in place one can rather easily prove the theorem. When considered as a map between curves $\phi - 1 : E(L) \to E(K)$ the kernel is $E(K)$, and the cardinality of the kernel is in the case of separable maps simply the degree of the map. It would require finding a differential on the elliptic curve and then proving some more theorems about separability and maps of differentials to show that the map $\phi - 1$ is in fact separable. Moreover the degree of ϕ is known to be p. As a result the number of points in the elliptic curve is close to q. It varies from q only by the affect of the -1 on the degree of the map ϕ. The degree map can be viewed as a quadratic form on the group of all isogenies, and therefore it satisfies a Cauchy-Schwarz inequality. Specifically

$$d(\phi - 1) - d(\phi) - d(1) \leq 2\sqrt{d(\phi)d(-1)} = 2\sqrt(q)$$

For those interested in looking further into this material, the approach given above follows that of Silverman[6] in his graduate text on elliptic curves. Joe Harris's book on Algebraic Geometry[2] is helpful as a basic introduction to those unfamiliar with Algebraic Geometry, and Emil Artin's classic text on Galois theory[1] is a good reference (although it does not discuss seperability very well). The three remaining books in the references by McKean, Ireland and Hida provide an interesting perspective on the various ways that elliptic curves are approached although they had little to do with this paper.

References

[1] Emil Artin, *Galois theory*, Dover Publications, 1998.

[2] Joe Harris, *Algebraic geometry*, Graduate Texts in Mathematics, vol. 133, Springer-Verlag, 1992.

[3] Haruzo Hida, *Geometric modular forms and elliptic curves*, World Scientific, Singapore, 2000.

[4] Kenneth Ireland and Michael Rosen, *A classical introduction to modern number theory*, Springer-Verlag, 1982.

[5] Henry McKean and Victor Moll, *Elliptic curves*, Cambridge University Press, 1997.

[6] Joseph H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 106, Sringer-Verlag, 1986.