2 Elliptic curves as abelian groups

In Lecture 1 we defined an elliptic curve as a smooth projective curve of genus 1 with a distinguished rational point. An equivalent definition is that an elliptic curve is an abelian variety of dimension one. An *abelian variety* is a smooth projective variety equipped with a group structure defined by rational maps (we will make this definition more precise below). Remarkably, the fact that we are working with projective varieties rather than affine varieties forces the group operation to be commutative, which is why they are called abelian varieties.

In this lecture we will prove that elliptic curves are abelian varieties by explicitly deriving the rational maps that define the group law. In the course of doing so we will verify that they do in fact satisfy the axioms required of a group operation.

2.1 The group law for Weierstrass curves

Recall from Lecture 1 that the group law for an elliptic curve defined by a Weierstrass equation is given by the following rule:

Three points on a line sum to zero, which is the point at infinity.

For convenience let us assume we are working over a field k whose characteristic is not 2 or 3. In this case we may assume that we are working with an elliptic curve E/k defined by a short Weierstrass equation

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

The case of a general Weierstrass equation $y^2+a_1xy+a_3y = x^3+a_2x^2+a_4x+a_6$ is essentially the same, but the formulas are slightly more complicated; see [8, III.2.3] for details and a proof that every elliptic curve can be defined by a Weierstrass equation.

Recall that although we typically specify our curves using an affine equation in the variables x and y, we are really working with the corresponding projective curve, which in this case is given by the homogeneous equation

$$E : y^2z = x^3 + Axz^2 + Bz^3.$$

In order to specify an elliptic curve we need not only an equation defining the curve, but also a distinguished rational point, which acts as the identity of the group. For curves in Weierstrass form we always take the point $O := (0 : 1 : 0)$ at infinity as our distinguished point; this is the unique point on the curve E that lies on the line $z = 0$ at infinity: if $z = 0$ then $x = 0$ and we may assume $y = 1$ after scaling the projective point $(0 : y : 0)$ by $1/y$; note that $x = z = 0$ forces $y \neq 0$, since $(0 : 0 : 0)$ is (by definition) not a projective point.

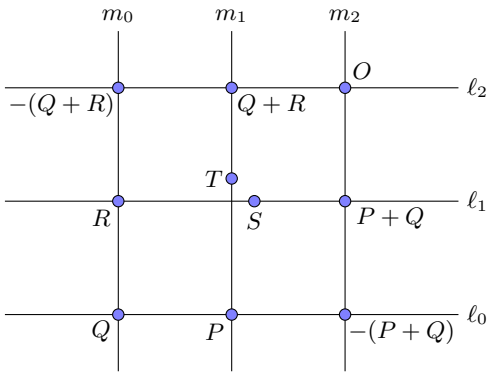
Every point $P \neq O$ on the curve E thus has a nonzero z-coordinate which we can scale to be 1, and we use the notation $P = (x_0, y_0) := (x_0 : y_0 : 1)$ to denote these affine points. Notice that the point $Q = (x_0, -y_0)$ also lies on the curve E, and the projective line through P and Q is defined by $x = x_0z$, which also passes through $O = (0 : 1 : 0)$. The three points P, Q, O lie on a line, so $P + Q + O = P + Q = O$, and therefore $Q = -P$.

Commutativity of the group law follows immediately from our definition, but associativity is not obvious. We will give two proofs. The first is geometric and depends on a genericity assumption that does not apply in special cases, but the proof is short and it provides some intuition as to *why* the group law is associative. The second is algebraic and handles all cases over any field whose characteristic is not 2 or 3, but we will rely on the computer algebra system Sage to do the heavy lifting. A formally verified proof of associativity that works in all characteristics is now available in the Lean `mathlib` library [1, 6].

2.1.1 A geometric proof of associativity in the generic case

This is an adaptation of the proof in [4, p. 28]. Let P, Q, R be points on an elliptic curve E over a field k that we may assume is algebraically closed (if the group law is associative over $\bar{k}$ then it is certainly also associative when we restrict to k). We shall also assume that P, Q, R, and the zero point O are all in *general position*. This means that in the diagram below there are no relationships among the points other than those that necessarily exist by construction; in particular the eight points $P, Q, R, O, \pm(P+Q), \pm(Q+R)$ are distinct and no three are collinear.

The line ℓ_0 through P and Q meets the curve E at a third point, $-(P+Q)$, and the line m_2 through O and $-(P+Q)$ meets E at $P+Q$. Similarly, the line m_0 through Q and R meets E at $-(Q+R)$, and the line ℓ_2 through O and $-(Q+R)$ meets E at $Q+R$. Let S be the third point where the line ℓ_1 through $P+Q$ and R meets E, and let T be the third point where the line m_1 through P and $Q+R$ meets E. See the diagram below:



We have $S = -((P+Q)+R)$ and $T = -(P+(Q+R))$. It suffices to show $S = T$. Suppose not. Let $g(x,y,z)$ be the cubic polynomial formed by the product of the lines ℓ_0, ℓ_1, ℓ_2 in homogeneous coordinates, and similarly let $h(x,y,z) = m_0 m_1 m_2$. We claim $g(T) \neq 0$. Indeed, if $g(T) = 0$ then T must lie on ℓ_1, since if it lies on ℓ_0 then so does $Q+R$ which is then collinear with P and Q, contradicting our general position assumption, and if it lies on ℓ_2 then so does P which is then collinear with O and $Q+R$, another contradiction; if T lies on ℓ_1 it must be equal to S, contrary to our supposition, because it cannot be equal to R or $P+Q$ (since neither is collinear with P and $Q+R$), and there are only three points in the intersection of ℓ_1 with E (by Bézout's theorem). Similarly, $h(S) \neq 0$.

It follows that g and h are linearly independent elements of the k-vector space V of homogeneous cubic polynomials in $k[x,y,z]$. The space V has dimension $\binom{3+2}{2} = 10$, thus the subspace of homogeneous cubic polynomials that vanish at the eight distinct points O, P, Q, R, $\pm(P+Q)$, and $\pm(Q+R)$ has dimension 2 and is spanned by g and h. The polynomial $f(x,y,z) = x^3 + Axz^2 + Bz^3 - zy^2$ that defines E is a nonzero element of this subspace, so we may write $f = ag + bh$ as a linear combination of g and h. Now $f(S) = f(T) = 0$, since S and T are both points on E, but $g(S), h(T) = 0$ and $g(T), h(S) \neq 0$, which implies that a and b are both zero, but this is a contradiction because f is not the zero polynomial.

This completes our geometric proof of associativity (in the generic case). In order to give a more general algebraic proof, and to be able to actually perform group operations explicitly, we need explicit formulas for computing the sum of two points.

2.2 The group law in algebraic terms

Let P and Q be two points on our elliptic curve $E\colon y^2 = x^3 + Ax + B$. We want to compute the point $R = P + Q$ by expressing the coordinates of R as rational functions of the coordinates of P and Q. If either P or Q is the point O at infinity, then R is simply the other point, so we assume that P and Q are affine points $P = (x_1, y_1)$ and $Q = (x_2, y_2)$. There are two cases.

Case 1. $x_1 \neq x_2$. The line $\overline{PQ}$ has slope $m = (y_2 - y_1)/(x_2 - x_1)$, which yields the linear equation $y - y_1 = m(x - x_1)$ for $\overline{PQ}$. This line is not vertical, so it intersects the curve E in a third affine point $-R = (x_3, -y_3)$. Plugging the equation for the line $\overline{PQ}$ into the equation for the curve E yields

$$(m(x - x_1) + y_1)^2 = x^3 + Ax + B.$$

Expanding the LHS and moving every term to the RHS yields a cubic equation

$$g(x) := x^3 - m^2x^2 + \cdots = 0,$$

where the ellipsis hides lower order terms in x. The monic cubic polynomial $g(x)$ has two roots $x_1, x_2 \in k$ and therefore factors in $k[x]$ as

$$g(x) = (x - x_1)(x - x_2)(x - x_3),$$

where $x_3 \in k$ is the x-coordinate of the third point $-R$ on the intersection of $\overline{PQ}$ and E. Comparing the coefficient of x^2 in the two expressions for $g(x)$ shows that $x_1 + x_2 + x_3 = m^2$, and therefore $x_3 = m^2 - x_1 - x_2$. We can then compute the y-coordinate $-y_3$ of $-R$ by plugging this expression for x_3 into the equation for $\overline{PQ}$, and we have

$$\begin{aligned} m &= (y_2 - y_1)/(x_2 - x_1), \\ x_3 &= m^2 - x_1 - x_2, \\ y_3 &= m(x_1 - x_3) - y_1, \end{aligned}$$

which expresses the coordinates of $R = P + Q$ as rational functions of the coordinates of P and Q as desired. To compute $P + Q = R$, we need to perform three multiplications (one of which is squaring m) and one inversion in the field k. We'll denote this cost $3\mathbf{M}+\mathbf{I}$; we are ignoring the cost of additions and subtractions because these are typically negligible compared to the cost of multiplications and (especially) inversions.

Case 2. $x_1 = x_2$. We must have $y_1 = \pm y_2$. If $y_1 = -y_2$ then $Q = -P$ and $P + Q = R = 0$. Otherwise $P = Q$ and $R = 2P$, and the line $\overline{PQ}$ is the tangent at P on the equation for E, whose slope we can compute by implicit differentiation. This yields

$$2y\,dy = 3x^2dx + A\,dx,$$

so at the point $P = (x_1, y_1)$ the slope of the tangent line is

$$m = \frac{dy}{dx} = \frac{3x_1^2 + A}{2y_1},$$

and once we know m we can compute x_3 and y_3 as above. Note that we require an extra multiplication (a squaring) to compute m, so computing $R = 2P$ has a cost of $4\mathbf{M}+\mathbf{I}$.

Remark 2.1. You might object that we have not formally defined implicit differentiation over an arbitrary field, nor have we shown that this gives us the slope of the tangent line. One can rigorously justify this (using Kähler differentials, for example), but it is easy to verify that it works in our case: if you plug $y = m(x - x_1) + y_1$ into the curve equation $E\colon y^2 = x^3 + Ax + B$ using the slope $m = (3x_1^2 + A)/2y_1$ we computed using implicit differentiation, you will find that x_1 is a double root, and since the point $(x_1, -y_1)$ does not lie on the line $L\colon y = m(x - x_1) + y_1$ unless $y_1 = 0$, the point (x_1, y_1) has multiplicity 2 in the intersection $E \cap L$, which implies that L is tangent to E at (x_1, y_1) as claimed.

With these equations in hand, we can now prove associativity as a formal identity, treating $x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3, A, B$ as indeterminates subject to the three relations implied by the fact that P, Q, R lie on the curve E. See the Sage worksheet

Lecture 2 Proof of associativity

for details, which includes checking all the special cases.

The equations above can be converted to projective coordinates by replacing x_1, y_1, x_2, and y_2 with x_1/z_1, y_1/z_1, x_2/z_2, and y_2/z_2 respectively, and then writing the resulting expressions for x_3/z_3 and y_3/z_3 with a common denominator. When $P \neq Q$ we obtain

$$\begin{aligned}
x_3 &= (x_2z_1 - x_1z_2)((y_2z_1 - y_1z_2)^2z_1z_2 - (x_2z_1 - x_1z_2)^2(x_2z_1 + x_1z_2))\\
y_3 &= (y_2z_1 - y_1z_2)((x_2z_1 - x_1z_2)^2(x_2z_1 + 2x_1z_2) - (y_2z_1 - y_1z_2)^2z_1z_2) - (x_2z_1 - x_1z_2)^3y_1z_2\\
z_3 &= (x_2z_1 - x_1z_2)^3z_1z_2
\end{aligned}$$

and for $P = Q$ we obtain

$$\begin{aligned}
x_3 &= 2y_1z_1(A^2(z_1^2 + 3x_1^2)^2 - 8x_1y_1^2z_1)\\
y_3 &= A(z_1^2 + 3x_1^2)(12x_1y_1^2z_1 - A^2(z_1^2 + 3x_1^2)^2) - 8y_1^4z_1^2\\
z_3 &= (2y_1z_1)^3
\end{aligned}$$

These formulas are more complicated, but they have the advantage of avoiding inversions, which are more costly than multiplications (in a finite field of cryptographic size inversions may be 50 or even 100 times more expensive than multiplications). With careful reuse of common subexpressions these formulas lead to a cost of 12**M** for addition (of distinct points) and 14**M** for doubling.

2.3 Elliptic curves as abelian varieties

An abelian variety is a smooth projective variety G/k equipped with morphisms $\mu\colon G \times G \to G$ and $i\colon G \to G$ and a k-rational point O such that for every field extension K/k the set $G(K)$ of K-rational points has the structure of a group with composition law given by μ, inverses given by i, and O as the identity element.

We have not formally defined what it means to be a smooth projective variety, but we have defined smooth projective plane curves C: these are defined by a polynomial in $f \in k[x, y, z]$ that is irreducible in $\bar{k}[x, y, z]$ such that there is no point $P \in C(\bar{k})$ at which the three (formal) partial derivatives of f simultaneously vanish. This example of a smooth projective variety suffices for our present purpose, as it includes the case of an elliptic curve. For the morphism μ we can take the rational maps defined by the polynomial expressions

we derived above for x_3, y_3, z_3 in terms of the projective coordinates x_1, y_1, z_1 and x_2, y_2, z_2, and for the inverse morphism i we simply take the map $(x : y : z) \mapsto (x : -y : z)$.

In the case of elliptic curves, the group law is commutative by construction. In fact commutativity holds for all abelian varieties [7, §4.3], which justifies their nomenclature, even though it is not obviously implied by the definition; indeed, with affine algebraic groups, which are defined exactly as abelian varieties but with the underlying algebraic variety affine rather than projective, the group operation is typically not commutative.

Remark 2.2. We have shown that elliptic curves are abelian varieties of dimension one (curves are algebraic varieties of dimension one by definition, regardless of their genus). We have not shown that every abelian variety of dimension one is an elliptic curve, which is beyond the scope of this course, but this is indeed the case (abelian varieties of dimension one are smooth projective curves, but one needs to show that they have genus 1, and that the group operation on the abelian variety necessarily coincides with that induced by the elliptic curve group law when we take the identity element as our distinguished point).

2.4 Edwards curves

Various alternative models of elliptic curves other than Weierstrass equations have been proposed over the years; each leads to different formulas for the group law that are ultimately equivalent to the formulas for curves in Weierstrass form, after applying a suitable isomorphism, but which may be more efficient to compute or have other advantages.

We give just one example here, a particular form of an *Edwards curve* [2, 3, 5]. Let a be a non-square element of a field k whose characteristic is not 2. Then the equation

$$x^2 + y^2 = 1 + ax^2y^2 \tag{1}$$

defines an elliptic curve with distinguished point $(0, 1)$.

Remark 2.3. The plane projective curve defined by equation (1) has two singular points at infinity, violating our requirement that an elliptic curve be smooth. However, this plane curve can be *desingularized* by embedding it in $\mathbb{P}^3(k)$. The points at infinity are then no longer rational, and do not play a role in the group operation on $E(k)$, whose elements can all be uniquely represented as solutions (x, y) to equation (1) above.

If we define

$$w := (ax^2 - 1)y, \qquad X := \frac{-2(w-1)}{x^2}, \qquad Y := \frac{4(w-1) + 2(a+1)x^2}{x^3},$$

then for any solution (x_0, y_0) to (1) with $x_0 \neq 0$ we obtain an affine point (X_0, Y_0) on the elliptic curve E/k defined by the Weierstrass equation

$$Y^2 = (X - a - 1)(X^2 - 4a).$$

(this is not a short Weierstrass equation, since the coefficient of X^2 is not zero, but for $\mathrm{char}(k) \neq 3$ the substitution $X = X' + a + 1$ yields a short Weierstrass equation).

If we map the solution $(0, 1)$ to the point at infinity on E and the solution $(0, -1)$ to the point $(a+1, 0)$ on E we obtain a bijection between the set of k-rational solutions to (1) and $E(k)$ (and similarly for all field extensions K of k, even though a may be a square in K). It is straightforward to check that this is in fact a bijection: if two points (x_0, y_0) map to

the same value of $X_0 := X(x_0, y_0)$ they must be of the form $(\pm x_0, y_0)$, but then the values of $Y_0 := Y(\pm x_0, y_0)$ will differ in sign unless $x_0 = 0$, but $(0, 1)$ and $(0, -1)$ are distinguished by the fact that one is mapped to the point at infinity and the other is not.

It follows that we can use the group law on E (three points on a line to sum to zero) to give the k-rational solutions to (1) the structure of a group isomorphic to $E(k)$ (and similarly if we replace k with an extension field K). One can then work out explicit formulas for this group law in terms of coordinates on the Edwards curve (1). We shall omit the details of these derivations (which are best done using a computer algebra system) and simply present the final result, which is quite pleasing.

The formula for adding points (x_1, y_1) and (x_2, y_2) in $E(k)$ is

$$(x_1, y_1) + (x_2, y_2) = \left(\frac{x_1 y_2 + x_2 y_1}{1 + a x_1 x_2 y_1 y_2}, \frac{y_1 y_2 - x_1 x_2}{1 - a x_1 x_2 y_1 y_2} \right), \tag{2}$$

which implies that the inverse of (x_1, y_1) is $(-x_1, y_1)$. In contrast to the formulas for curves in Weierstrass form, the formula in (2) is well defined for every pair of points (x_1, y_1) and (x_2, y_2) in $E(k)$.

To prove this, let us suppose for the sake of obtaining a contradiction that one of the denominators in (2) is zero for some pair of inputs $(x_1, y_1), (x_2, y_2)$. Then we must have

$$(1 + a x_1 x_2 y_1 y_2)(1 - a x_1 x_2 y_1 y_2) = 1 - a^2 x_1^2 x_2^2 y_1^2 y_2^2 = 0,$$

so $a^2 x_1^2 x_2^2 y_1^2 y_2^2 = 1$, and therefore x_1, x_2, y_1, y_2 are all nonzero. Applying this and the curve equation (twice) yields

$$x_1^2 + y_1^2 = 1 + a x_1^2 y_1^2 = 1 + \frac{1}{a x_2^2 y_2^2} = \frac{x_2^2 + y_2^2}{a x_2^2 y_2^2}.$$

By adding or subtracting $2 x_1 y_1 = \pm 2/(a x_2 y_2)$ to both sides we can obtain

$$(x_1 \pm y_1)^2 = \frac{(x_2 \pm y_2)^2}{a x_2^2 y_2^2},$$

with either choice of sign on the LHS (the sign on the RHS may vary, but in any case the numerator of the RHS is a square). Since x_1 and y_1 are nonzero, one of $x_1 + y_1$ and $x_1 - y_1$ is nonzero, and this implies that a is a square in k, but this is a contradiction, since we assumed from the beginning that a is not a square in k.

Remark 2.4. The formula in (2) works over extension fields at all points where it is well defined, but it is only for extensions K/k where a is not a square that it is guaranteed to be well defined at every K-rational point (and if a is a square the desingularization of the projective curve defined by (1) will have two rational points at infinity not handled by (1)).

As written, the group law involves five multiplications and two inversions (ignoring the multiplication by a, which we can choose to be small), which is greater than the cost of the group operation in Weierstrass form. However, in projective coordinates we have

$$\frac{x_3}{z_3} = \frac{z_1 z_2 (x_1 y_2 + x_2 y_1)}{z_1^2 z_2^2 + a x_1 x_2 y_1 y_2}, \qquad \frac{y_3}{z_3} = \frac{z_1 z_2 (y_1 y_2 - x_1 x_2)}{z_1^2 z_2^2 - a x_1 x_2 y_1 y_2}.$$

There are a bunch of common subexpressions here, and in order to compute z_3, we need a common denominator. Let $r = z_1z_2$, let $s = x_1y_2 + x_2y_1$, let $t = ax_1y_2x_2y_1$, and let $u = y_1y_2 - x_1x_2$. We then have

$$x_3 = rs(r^2 - t), \qquad y_3 = ru(r^2 + t), \qquad z_3 = (r^2 + t)(r^2 - t).$$

This yields a cost of 12**M**. If we compute s as $s = (x_1 + y_1)(x_2 + y_2) - x_1x_2 - y_1y_2$, the cost is reduced to 11**M**.

A simple Sage implementation of these formulas can be found here:

Lecture 2 Group law on Edwards curves

Because the expression in (2) is well defined at every point in $E(k)$, we do not need separate formulas for addition and doubling.[1] Moreover, we don't even need to check the cases where one or both points is the identity element, or one is the negation of the other; the same formula works in every case. Such formulas are said to be *complete*, and they have two distinct advantages. First, they can be implemented very efficiently as a straight-line program with no branching. Second, they protect against what is known as a *side-channel* attack. If you are using different formulas for addition and doubling, it is possible that an adversary may be able to externally distinguish these cases, e.g. by monitoring the CPU (electronically, thermally, or even acoustically) and noticing the difference in the time required or energy used by each operation. They can then use this information to break a cryptosystem that performs scalar multiplication by an integer n that is meant to be secret (as in Diffie-Hellman key exchange, for example), because the sequence of doubling and1 adding used in scalar multiplication effectively encodes the binary representation of n. Using complete formulas prevents a side-channel attack because exactly the same sequence of instruction is executed for every group operation.

Having said that, if you know you want to double a point and are not concerned about a side-channel attack, there are several optimizations that can be made to the formulas above (these include replacing $1 + cx^2y^2$ with $x^2 + y^2$). This reduces the cost of doubling on an Edwards curve to 7**M**, half the 14**M** cost of doubling a point in Weierstrass coordinates [2].

The explicit formulas database contains optimized formulas for Edwards curves and various generalizations, as well as many other forms of elliptic curves. Operation counts and verification scripts are provided with each set of formulas.

We should note that, unlike Weierstrass equations, not every elliptic curve can be defined by an equation in Edwards form. In particular, an Edwards curve always has a rational point of order 4, the point $(1, 0)$, but most elliptic curves do not have a rational point of order 4.

References

[1] David Kurniadi Angdinata and Junyan Xu, *An elementary formal proof of the group law on Weierstrass elliptic curves in any characteristic*, arXiv:2302.10640, 2023.

[2] Daniel J. Bernstein and Tanja Lange, *Faster addition and doubling on elliptic curves*, Advances in Cryptology - ASIACRYPT 2007, Lecture Notes in Computer Science **4833**, Springer-Verlag, New York (2007), 29–50.

[1] See [3] for formulas that achieve this for every point in $E(\bar{k})$ without assuming a is a non-square, and also for the more general case of *twisted Edwards curves* $dx^2 + y^2 = 1 + ax^2y^2$.

[3] Daniel J. Bernstein and Tanja Lange, *A complete set of addition laws for incomplete Edwards curves*, Journal of Number Theory **131** (2011), 858–872.

[4] J. W. S. Cassels, *Lectures on elliptic curves*, London Mathematical Society Student Texts **24**, Cambridge University Press, 1991.

[5] Harold M. Edwards, *A normal form for elliptic curves*, Bulletin of the American Mathematical Society **44** (2007), 393–422.

[6] The Mathlib Community, *The Lean mathematical library*, available at `https://github.com/leanprover-community/mathlib4`.

[7] Igor R. Shafarevich, *Basic algebraic geometry*, Springer, 1994.

[8] Joseph H. Silverman, *The arithmetic of elliptic curves*, 2nd ed., Springer, 2009.