2.1 The elliptic curve group law

Recall from Lecture 1 the defining property of the group law for an elliptic curve defined by a Weierstrass equation $y^2 = x^3 + Ax + B$:

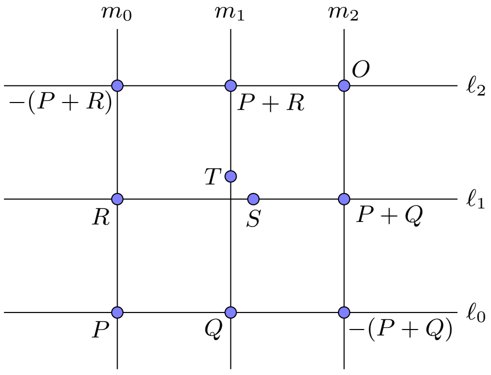
Three points on a line sum to zero, which is the point at infinity.

It is easy to determine the inverse of a point (just negate the y coordinate), and it is obvious that group operation is commutative. Associativity is not obvious, and while it can be rigorously proven algebraically, this is a tedious task that does not yield much insight. So we will give two proofs. The first will only apply to the generic case but it is short and provides some explanation as to *why* the group operation is associative. The second will be algebraic and fully rigorous, but we will let Sage do all the dirty work for us.

2.1.1 A geometric proof of associativity in the generic case

This is an adaptation of the proof in [2, p. 28]. Let P, Q, and R be three points on an elliptic curve $E(k)$ for some field k that we may assume is algebraically closed. 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).

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



We have $S = -(Q+P) + R$ and $T = -(Q + (P+R))$. 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 $h(x,y,z) = m_0m_1m_2$. We may assume $g(T) \neq 0$ and $h(S) \neq 0$, since the points are in general position and $S \neq T$. Thus 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 10, thus the subspace of homogeneous cubic polynomials that vanish at the eight points O, P, Q, R, $\pm(Q+P)$, and $\pm(P+R)$ 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. But $f(S) = f(T) = 0$, since S and T are both points on E, which implies that a and b are both zero. This contradicts the linear independence of g and h, since f is not the zero polynomial.

2.1.2 The group law in algebraic terms

Let $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$ be two points on E. We will compute the sum $P + Q = R = (x_3, y_3, z_3)$ by expressing the coordinates of R as rational functions of the coordinates of P and Q. If either P or Q is the point at infinity, then R is simply the other point, so we assume that P and Q are affine points with $z_1 = z_2 = 1$. 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 equation $y - y_1 = m(x - x_1)$. The point $-R = (x_3, -y_3, 1)$ is on this line, thus $-y_3 = m(x_3 - x_1) + y_1$. Substituting for y_3 in the Weierstrass equation for E yields

$$(m(x_3 - x_1) + y_1)^2 = x_3^3 + Ax_3 + B.$$

Simplifying, we obtain $0 = x_3^3 - m^2 x_3^2 + \cdots$, where the ellipsis hides lower order terms. The values x_1 and x_2 satisfy the same cubic equation, thus its roots are x_1, x_2, and x_3, and the sum of these roots is equal to the negation of the quadratic coefficient $-m^2$. Thus $x_3 = m^2 - x_1 - x_2$. To sum up, we have

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

Thus to compute $P + Q = R$, we need to perform three multiplications (one of which is a squaring) and one inversion in the field k. We'll denote this cost $3\mathbf{M}+\mathbf{I}$.

Case 2. $x_1 = x_2$. If $y_1 \neq y_2$, then they must be opposite points and $R = 0$. Otherwise $P = Q$, and we compute the slope of the tangent line by implicitly differentiating the Weierstrass equation for E. This yields $2y\,dy = 3x^2\,dx + A\,dx$, so $m = \frac{dy}{dx} = \frac{3x_1^2+A}{2y_1}$. The formulas for x_3 and y_3 are then the same as before. Note that we require an extra multiplication (a squaring) here, so computing $R = 2P$ has a cost of $4\mathbf{M}+\mathbf{I}$.

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 indeterminants subject to the three relations implied by the fact that P, Q, and R all lie on the curve E. See the Sage worksheet

18.783 Lecture 2: Proof of Associativity.sws

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. This has 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). This increases the number of multiplications to $12\mathbf{M}$ in case 1 (general addition) and $14\mathbf{M}$ in case 2 (doubling).

2.2 Edwards curves

There are many alternative representations of elliptic curves that have been proposed. We give just one example here, Edwards curves [1, 3], which have two significant advantages over Weierstrass equations. Let d be a non-square element of a field k (assumed to have characteristic not equal to 2, as usual). Then the equation

$$x^2 + y^2 = 1 + dx^2y^2$$

defines an elliptic curve with distinguished point $(0, 1)$.[1] The group operation is given by

$$(x_3, y_3) = \left(\frac{x_1y_2 + x_2y_1}{1 + dx_1x_2y_1y_2}, \frac{y_1y_2 - x_1x_2}{1 - dx_1x_2y_1y_2}\right).$$

As written, this involves five multiplications and two inversions (ignoring the multiplication by d, 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

$$\frac{x_3}{z_3} = \frac{z_1z_2(x_1y_2 + x_2y_1)}{z_1^2z_2^2 + dx_1x_2y_1y_2}, \qquad \frac{y_3}{z_3} = \frac{z_1z_2(y_1y_2 - x_1x_2)}{z_1^2z_2^2 - dx_1x_2y_1y_2}.$$

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 $t = dx_1y_2x_2y_1$, 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**.

The remarkable thing about these formulas is that they handle every case; there are not separate formulas for addition and doubling, and adding opposite points or the identity element works the same as the general case. Such formulas are called *complete*, and they have two distinct advantages. First, they can be implemented very efficiently because there is no branching. Second, they protect against what is known as a *side-channel* attack. If an adversary can distinguish whether you are doubling or adding points, e.g. by monitoring the CPU and noticing the difference in the time required by each operation, they can break a cryptosystem that performs scalar multiplication by an integer that is meant to be secret.

Having said that, if you know you are going to be doubling and are not concerned about a side-channel attack, there are several optimizations that can be made (these include replacing $1 + dx^2y^2$ with $x^2 + y^2$). This reduces the cost of doubling a point on an Edwards curves to 7**M**, which is a huge improvement over the 14**M** cost of doubling a point in Weierstrass coordinates.

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 put into Edwards form. In particular, an Edwards curve always has a rational point of order 4, the point $(1, 0)$, but this is not true of many elliptic curves.

[1] Technical point: there are two points at infinity, both of which are singular, 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 $E(k)$.

References

[1] D. Bernstein and T. 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.

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

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