1.8 26/06/25 - The Geometry of Cubic Curves II: Proof Sketch of Associativity

Today, our goal will be to finish the proof sketch of 1.7.7. To do this, I will outline the proof of the associativity of the group law on the cubic using the circle of ideas involving divisors, genus, and the Riemann-Roch Theorem (although I will not use this language).

But before that, I want to take a second to restate the Mordell-Weil theorem in this language as

Theorem 1.8.1 (Mordell-Weil Theorem)**.** Let K be a number field (e.g., $K = \mathbb{Q}$). Then for any smooth cubic curve $C \subset \mathbb{P}^2_K$ with an $\mathscr{O} \in C(K)$, the abelian group $C(K)$ described in 1.7.7 is finitely generated.

Make sure you understand that this is saying the same thing as: there are finitely many "seed" points on $C(K)$ such that every other point can be obtained from these by iteration of the "chord and tangent processes". One of our goals for the summer will be to prove (a simple case of) 1.8.1.

Unimportant Remark 1.8.2. Note that the theorem 1.8.1 is specific to number fields (or more generally global fields). It is not true if we take something like $K = \mathbb{C}$ (an uncountable group has a hard time being finitely generated!), and it is obviously true over a finite field $K = \mathbb{F}_q$, since then $C(\mathbb{F}_q)$ is just finite! This theorem is less obvious but true over fields of the form $K = \mathbb{F}_q(t)$, which are also global fields.

Given this discussion, let's move on to

1.8.A Proof Sketch of Associativity

To prove associativity of the group law, we need to study what it really means to say something like $S = P \oplus Q$. Specifically, we will give an equivalent definition of the group law $\oplus$ on $C(K)$ (1.8.8)–modulo two blackboxes (1.8.11 and 1.8.14), which you can fill in later–and see that this definition is visibly associative. (The following argument is an expanded version of one of the two found in [1, §7].)

Throughout this section, we will work with the following example to see ideas in practice: let K be a field of characteristic other than 2 or 3, and let $C := \mathbb{V}(Y^2Z - X(X+Z)(X+4Z))$ with $\mathscr{O} := [0:1:0]$.

Exercise 1.8.3. Check that when $\operatorname{ch} K \neq 2, 3$, the cubic curve $C = \mathbb{V}(Y^2Z - X(X+Z)(X+4Z)) \subset \mathbb{P}^2_K$ is smooth.

We'll take the points $P = [-4:0:1]$ and $Q = [-2:2:1]$, so that $U := P * Q = [2:6:1]$ and $S := P \oplus Q = [2:-6:1]$. The affine patch $Z \neq 0$ on this curve is pictured in 1.8.4.

Recall that if A is a homogeneous polynomial in $K[X,Y,Z]$, then it doesn't quite make sense to "evaluate" A at points of $\mathbb{P}^2(K)$. However, if A and B are both homogeneous polynomials of the same degree, then it at least makes sense to evaluate A/B at points of $\mathbb{P}^2(K)$ where B doesn't vanish. We've done this before already when we've taken $A = X, Y$ and $B = Z$; this is what gives us the coordinates $x = X/Z$ and $y = Y/Z$.

For another example, in the curve of 1.8.4, we could take A to be a linear form defining the line L_{PQ}, say $A := Y - X - 4Z$, and similarly B to be a linear form defining the vertical line $L_{U\mathscr{O}}$, say $B := X - 2Z$. Then the quotient would be

$$f := \frac{A}{B} = \frac{Y - X - 4Z}{X - 2Z} = \frac{y - x - 4}{x - 2} = \frac{1 - t - 4s}{t - 2s}, \tag{1.8.5}$$

where as before $(x, y) = (X/Z, Y/Z)$, and $(t, s) = (X/Y, Z/Y)$. It now does make sense to evaluate this fraction 1.8.5 away from the locus where $B = 0$. In particular, we can evaluate this function on our curve C, away from the three intersection points $\mathscr{O}, S, U$. For instance, what is its value at $(x, y) = (0, -1)$? It is $5/2$.

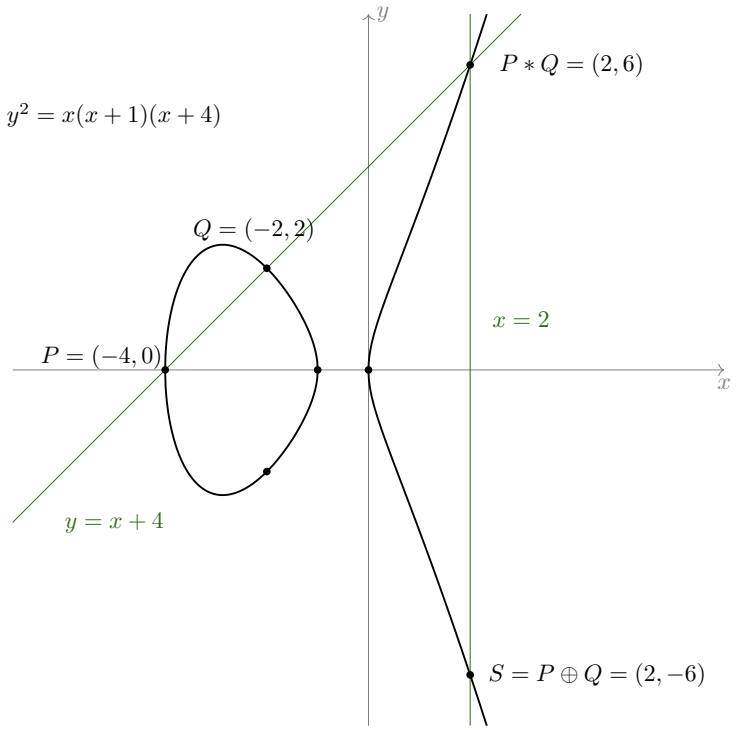

Figure 1.8.4: The $\{Z \neq 0\}$ points of the projective curve $\mathbb{V}(Y^2Z - X(X+Z)(X+4Z)) \subset \mathbb{P}^2_K$.

Where does this function vanish on our curve? Of course at the points P, Q, and U–the locus of intersection of $\mathbb{V}(A)$ and C. Where is it not defined, or where does it blow-up, or where is it infinite? Of course, at the points $\mathscr{O}, S, U$–the locus of intersection of $\mathbb{V}(B)$ and C. But wait: U belongs to both of these. What is happening there?

Exercise 1.8.6. (This exercise is for those who know a little calculus.) Show that if $K = \mathbb{R}$, then

$$\lim_{x \to 2} \frac{\sqrt{x(x+1)(x+4)} - x - 4}{x - 2}$$

exists and is nonzero by calculating it.

In fact, for any field K, the "zero" coming from A and the "pole" coming from B exactly "cancel out" at U, leaving us with no zeroes or poles at this point. Therefore, we have produced an *algebraic function* $f = A/B$ on C with the property that it has zeroes at P and Q, and poles at $\mathscr{O}$ and S.

Exercise 1.8.7. If you know what these words mean, show that the morphism $C \smallsetminus \{U\} \to \mathbb{P}^1_K$ defined by $[A : B]$ extends to give a morphism $C \to \mathbb{P}^1_K$ which maps U to a point different than 0 and ∞. (This is an instance of the "curve-to-projective" extension theorem.) If you don't know what these words mean, you may either try to interpret them or ignore this exercise.

We could take this to be a *definition* of $S = P \oplus Q$:

Definition 1.8.8. In the above set-up, given $P, Q, S \in C(K)$, we define $P \oplus Q = S$ iff there is a nonzero algebraic function f on C with zeroes at P and Q and poles at $\mathscr{O}$ and S (and no zeroes and poles elsewhere).

Of course, for this to be a definition of $P \oplus Q$, we need to do three things:

(a) we need to define what it means to be an algebraic function on the curve C, and what zeroes and poles of an algebraic function mean,

(b) we need to show that given any P and Q, such an S exists, and
(c) we need to show that given any P and Q, such an S is unique.

Let's deal with these issues one at a time.

Given an integral curve $C \subset \mathbb{P}^2_K$, we define the *field of algebraic functions* on C, denoted $K(C)$, to be the field consisting of functions of the form 1.8.5. To say more, recall what it means to say for C to be integral: it means that any defining equation $F \in K[X,Y,Z]$ for C is irreducible. Since $K[X,Y,Z]$ is a UFD, this is the same as saying that F is prime, or that $K[X,Y,Z]/(F)$ is an integral domain. Therefore, we may consider its fraction field, denoted

$$K(\mathrm{Cone}(C)) := \mathrm{Frac}\, K[X,Y,Z]/(F).$$

Now this field is too big and not what we're looking for: it contains things like X, which don't make sense as functions on C. (The notation is suggesting what this is: it is the field of algebraic functions on the cone over C, which is $\mathrm{Cone}(C) := \mathbb{V}(F) \subset \mathbb{A}^3_K$.) What we need is a smaller subfield $K(C) \subset K(\mathrm{Cone}(C))$, which consists of all fractions of the form $[A]/[B]$, where A and B can be chosen to be homogeneous of the same degree.

Definition 1.8.9. Let K be a field and $C \subset \mathbb{P}^2_K$ be an integral projective curve with defining equation $F \in K[X,Y,Z]$. Then we define the *algebraic function field* $K(C)$ of C to be

$$K(C) := \left\{ \frac{[A]}{[B]} : \begin{array}{l} [A],[B] \in K[X,Y,Z]/(F), \text{ such that } [B] \neq 0 \text{ and} \\ A \text{ and } B \text{ can be chosen to be homogeneous of the same degree} \end{array} \right\} \subset K(\mathrm{Cone}(C)).$$

The objects of $K(C)$ can then be thought of as genuine K-valued functions on $C(K)$, except possibly for finitely many points where they have the value "∞".

Example 1.8.10. Let $K = \mathbb{C}$ and $C = \mathbb{V}(Y^2Z - X(X+Z)(X+4Z)) \subset \mathbb{P}^2_{\mathbb{C}}$. Then $\mathbb{C}(C)$ contains things like $x := [X]/[Z]$, $y := [Y]/[Z]$, $t := [X]/[Y]$, and $s := [Z]/[Y]$. The class $[X]/[1]$ belongs to $\mathbb{C}(\mathrm{Cone}(C))$ but not to $\mathbb{C}(C)$. The fraction

$$\frac{[X^3 - Z^3]}{[Y^2Z - X^3 - Z^3]} = \frac{x^3 - 1}{y^2 - x^3 - 1} = \frac{t^3 - s^3}{s - t^3 - s^3}$$

defines an element of $\mathbb{C}(C)$, but the fraction

$$\frac{[X^3 - Z^3]}{[Y^2Z - X(X+Z)(X+4Z)]}$$

does not, since the denominator is zero. As elements of $\mathbb{C}(C)$, we have that

$$y^2 = x(x+1)(x+4) \text{ and } s = t(t+s)(t+4s).$$

In fact,

$$\mathbb{C}(C) \cong \mathrm{Frac}\, \mathbb{C}[x,y]/(y^2 - x(x+1)(x+4)) \cong \mathrm{Frac}\, \mathbb{C}[t,s]/(s - t(t+s)(t+4s)).$$

More generally, the function fields of a curve is the same as the "function field" of any of its nonempty affine patches (2.3.2).

Now we're going to need one important fact:

Fact 1.8.11. Let $C \subset \mathbb{P}^2_K$ be a smooth integral[a] curve with algebraic function field $K(C)$.

(a) For each $P \in C(K)$, we have a (discrete) valuation

$$v_P : K(C)^* \to \mathbb{Z}$$

which measures the order of vanishing of an algebraic function at P.
(b) Given an $f \in K(C)^*$ and a $P \in C(K)$, we say that P is a *zero* (resp. *pole*) of f of order $v_P(f)$ (resp. $-v_P(f)$) if $v_P(f) > 0$ (resp. $v_P(f) < 0$). A zero (resp. pole) of order one is said to be a simple zero (resp. simple pole), and similarly for double/triple zeroes/poles, etc.

(c) If we have a function f of the form $f = [A]/[B]$ where $A, B \in K[X, Y, Z]$ are homogeneous of the same degree and $[B] \neq 0$, then for any $P \in C(K)$, if $A(P) = 0$ but $B(P) \neq 0$, then P is a zero of f. Similarly, if $A(P) \neq 0$ but $B(P) = 0$, then P is a pole of f. (More care is needed if $A(P) = B(P) = 0$; c.f. 1.8.6).

(d) Moreover, if A and B are linear forms, $A(P) = 0$ but $B(P) \neq 0$, and the line $\mathbb{V}(A)$ is *not* tangent to C at P, then $v_P(f) = 1$, i.e., f has a simple zero at P. Similarly, if $A(P) \neq 0$ and $B(P) = 0$, and the line $\mathbb{V}(B)$ is not tangent to C at P, then $v_P(f) = -1$, i.e., f has a simple pole at P. More generally, if $A(P) = 0$ but $B(P) \neq 0$, and $L = \mathbb{V}(A)$, then $v_P(f) = m_P(C, L)$ is the intersection multiplicity of C and L at P as in the proof of 1.7.1.

[a] See also 2.3.3.

For a proof of the first part of 1.8.11, see [8, §II.1] and the references cited there, where v_P is denoted by ord_P. For a proof of the second part (unfortunately not using exactly the same language), see [6, Theorem 1, §3.2].

Remark 1.8.12. The operation v_P on the field $K(C)$ behaves very similarly to the p-adic valuation v_p on the field $\mathbb{Q}$ of rational numbers. This suggests that maybe we can interpret the p-adic valuation in this language. For instance, where does the "function" $12/5$ have zeroes and poles? Of course, it has a "double zero" at 2, a "simple zero" at 3, and a "simple pole" at 5. On what geometric object are rational numbers rational functions?

Unimportant Remark 1.8.13. The same as in 1.8.11 can be done for other closed points $P \in C$, possibly with bigger residue fields. (E.g., this might correspond to something like the $(t^2 + 1)$-adic valuation on the field $\mathbb{F}_3(t)$.) If you have no idea what this means, you can safely ignore this bit.

At this point, we can at least make sense of the individual components of 1.8.8: given a field K, a smooth cubic $C \subset \mathbb{P}^2_K$ with $\mathscr{O} \in C(K)$, and given points $P, Q, S \in C(K)$, we say that $P \oplus Q = S$ iff there is an $f \in K(C)^\times$ such that $v_P(f) = v_Q(f) = 1$ and $v_{\mathscr{O}}(f) = v_S(f) = -1$ (when $P, Q, \mathscr{O}, S$ are all distinct; otherwise, say, if $P = Q$, then we require that $v_P(f) = 2$, etc.) and such that f has no zeroes or poles elsewhere (interpreted correctly, see 1.8.15). To show that this is a definition of $P \oplus Q$, we still need to show that such an S exists and is unique. The existence part, however, is OK: we have seen above that the point $S = P \oplus Q$ defined by our "chord and tangent process" works. (The multiplicities are automatically taken care of by various parts of 1.8.11; details omitted for now.) The uniqueness, however, needs another important fact, namely

Fact 1.8.14. Let K be a field, $C \subset \mathbb{P}^2_K$ a smooth integral curve with algebraic function field $K(C)$. If $S_1, S_2 \in C(K)$ are two distinct points, then there does not exist an $f \in K(C)^\times$ with a simple zero at S_1, a simple pole at S_2, and no zeroes or poles elsewhere, i.e., with $v_{S_1}(f) = -v_{S_2}(f) = 1$, and $v_S(f) = 0$ for all $S \neq S_1, S_2$.

Unimportant Remark 1.8.15. Here again, "elsewhere" needs to be interpreted correctly as: at no other closed points (c.f. 1.8.13). This is not a problem if we assume K is algebraically closed, which we can certainly do for the purposes of proving 1.7.7: simply pass to an algebraic closure $\overline{K}$ and prove the result there, and then note that $C(K) \subset C(\overline{K})$ forms a subgroup. The reason I wanted to not avoid this approach from the outset is that I wanted to emphasize that the function field $K(C)$ can be defined even when K is not algebraically closed; some authors (such as Silverman in [8]) decide to work almost exclusively with $\overline{K}(C)$ instead, where $\overline{K}$ denotes an algebraic closure of K.

This fact (1.8.14) is something special to plane cubic curves (or actually smooth plane curves of degree ≥ 3):

Exercise 1.8.16. Let K be a field, $C \subset \mathbb{P}^2_K$ be a line or a smooth conic. Show that for any two distinct points $S_1, S_2 \in C(K)$, there is an $f \in K(C)^\times$ with $v_{S_1}(f) = -v_{S_2}(f) = 1$, and $v_S(f) = 0$ for all $S \neq S_1, S_2$. (Hint: Use a nice coordinate system. See also 1.6.4.)

For the cognoscenti, the rough idea behind the proof of 1.8.14 is as follows. If f is such a function, then f induces a morphism $C \to \mathbb{P}^1_K$ (c.f. 1.8.7). If f has the mentioned properties, then f has degree 1, and is hence necessarily an isomorphism. But C and $\mathbb{P}^1_K$ are *not* isomorphic: they have different

genera, namely $g(C) = 1$ but $g(\mathbb{P}^1_K) = 0$. The notion of genus is easiest to understand when $K = \mathbb{C}$, where it just means "the number of holes". However, it also makes sense in positive characteristic, with one definition being: it is the integer that makes the Riemann-Roch Theorem true. For the notion of genus, see for instance [6, §8.3]. For the Riemann-Roch Theorem, see [6, §8.6]. To see how 1.8.14 follows from the Riemann-Roch Theorem, see [8, Lemma III.3.3].

At this point, uniqueness is disposed of by

Lemma 1.8.17. Let K be a field, $C \subset \mathbb{P}^2_K$ a smooth cubic curve with $\mathscr{O} \in C(K)$. Suppose we are given points $P, Q, S_1, S_2 \in C(K)$, and assume for simplicity that $P \neq Q$ and $S_1, S_2 \neq \mathscr{O}$. Suppose for $i = 1, 2$ that there are $f_i \in K(C)^\times$ such that $v_P(f_i) = v_Q(f_i) = 1$ and $v_{\mathscr{O}}(f_i) = v_{S_i}(f_i) = -1$, with both f_1 and f_2 having no zeroes or poles elswhere. Then $S_1 = S_2$.

Proof. If $S_1 \neq S_2$, then $f := f_2/f_1$ has the property that $v_{S_1}(f) = 1$ and $v_{S_2}(f) = -1$, and that $v_S(f) \neq 0$ elsewhere (1.8.15); this contradicts 1.8.14. ∎

The cases where $P = Q$ or one of the S_i coincide with $\mathscr{O}$ is similar and left to the reader:

Exercise 1.8.18. Show that the result of 1.8.17 is also true if $P = Q$ or if one of the S_i is $\mathscr{O}$. For instance, show that if $P = Q$ but $S_1, S_2 \neq \mathscr{O}$, and for $i = 1, 2$ there are $f_i \in K(C)^\times$ such that $v_P(f_i) = 2$ and $v_{\mathscr{O}}(f_i) = v_{S_i}(f_i) = -1$, with both f_1 and f_2 having no zeroes or poles elsewhere, then $S_1 = S_2$. State and prove a similar result when (a) $P \neq Q$, but one of the S_i coincides with $\mathscr{O}$, and (b) $P = Q$ and one of the S_i coincides with $\mathscr{O}$. (Hint: The same proof works.)

Unimportant Remark 1.8.19. The above exercise 1.8.18 shows that it would be very convenient to have the language of adding and subtracting zeroes and poles as abstract objects; this would allow us to give a uniform proof of 1.8.17 and the result in 1.8.18, which clearly contain the same ideas. This is indeed possible, and is achieved by introducing the notion of a *divisor* on a curve, which is basically an integer-linear combination of points on it. If you're interested in this notion, come and talk to me, or see [6, §8.1] or [8, §II.3].

Finally, at this stage, 1.8.8 makes sense: we've defined what the terms mean, and shown that such an S exists and is unique. The associativity statement now follows from

Lemma 1.8.20. Let K be a field, $C \subset \mathbb{P}^2_K$ be an (integral) smooth cubic curve, and let $\mathscr{O} \in C(K)$. Suppose we are given points $P, Q, R, T \in C(K)$, and let $S = P \oplus Q$. Suppose for simplicity that $\mathscr{O}, P, Q, R, S, T$ are all distinct. Then we have that

$$T = (P \oplus Q) \oplus R$$

if and only if there is an $h \in K(C)^\times$ which has simple zeroes at P, Q, R, a double pole at $\mathscr{O}$, and a double pole at T, and no other zeroes or poles, i.e.,

$$v_P(h) = v_Q(h) = v_R(h) = 1 \text{ and } v_{\mathscr{O}}(h) = -2 \text{ and } v_T(h) = -1,$$

and such that $v_U(h) = 0$ for all $U \neq \mathscr{O}, P, Q, R, T$.

Proof. We need to show two implications. In both cases, from the equality $S = P \oplus Q$ and 1.8.8, there is an $f \in K(C)^\times$ such that $v_P(f) = v_Q(f) = -v_{\mathscr{O}}(f) = -v_S(f) = 1$, and which has no other poles or zeroes.

(a) Suppose first that $T = (P \oplus Q) \oplus R$, and we have to produce an h with the required properties. From $T = S \oplus R$, there is a $g \in K(C)^\times$ such that $v_S(g) = v_R(g) = -v_{\mathscr{O}}(g) = -v_T(g) = 1$, and such that g has no other poles or zeroes. Then taking $h = f \cdot g$ works (check!).

(b) Conversely, suppose that there is an h with the required properties, and we have to show that $T = S \oplus R$. Well, by 1.8.8, it suffices to produce a $g \in K(C)^\times$ such that $v_S(g) = v_R(g) = -v_{\mathscr{O}}(g) = -v_T(g) = 1$, and such that g has no other poles or zeroes. But the choice $g := h/f$ works (check!).

■

Again, edge cases are left as an

Exercise 1.8.21.

(a) Show that an appropriate version of the result of 1.8.20 also holds when some of the $\mathscr{O}, P, Q, R, S, T$ coincide.
(b) Show that associativity of $\oplus$ follows from 1.8.20 and the edge cases covered in (a).

This completes our proof sketch of the associativity of the group law on a smooth cubic with a rational point.