1.9 26/06/29 - Weierstrass Equations I

Today, we will roughly follow §1.3-1.4 of [2].

Last time, we showed that if K is a field and $C \subset \mathbb{P}^2_K$ a smooth cubic curve, then for any $\mathscr{O} \in C(K)$, the operation $\oplus$ given by $C(K) \ni P, Q \mapsto P \oplus Q := \mathscr{O} * (P * Q)$ makes $C(K)$ into an abelian group. To finish up that discussion, I only want to remark one more thing: a priori, this group structure depends on the choice of $\mathscr{O}$, so it's not clear what happens if you take two different points $\mathscr{O}_1, \mathscr{O}_2 \in C(K)$. Are the resulting group structures on $C(K)$ the same? They can't *literally* be the same if $\mathscr{O}_1 \neq \mathscr{O}_2$, since they have different identity elements. Is there a different way they are the same? This I leave you to explore in Exercise 2.3.9.

At this point, we are finally in a position to be able to define the eponymous *elliptic curves*.

1.9.A Elliptic Curves and Weierstrass Equations

Let me start by giving two different definitions of elliptic curves.

Definition 1.9.1. Let K be a field. An *elliptic curve* over K is a pair $(E, \mathscr{O})$, where E is a smooth projective geometrically integral curve of genus one, and $\mathscr{O} \in E(K)$.

Definition 1.9.2. Let K be a field. An *elliptic curve* over K is a curve $E \subset \mathbb{P}^2_K$ with equation of the form

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3, \tag{1.9.3}$$

where $a_1, a_2, a_3, a_4, a_6 \in K$ are such that $\Delta \neq 0$, where Δ is given by the formula

$$\begin{aligned}\Delta = &-a_1^6a_6 + a_1^5a_3a_4 - a_1^4a_2a_3^2 - 12a_1^4a_2a_6 + a_1^4a_4^2 + 8a_1^3a_2a_3a_4 + a_1^3a_3^3\\ &+ 36a_1^3a_3a_6 - 8a_1^2a_2^2a_3^2 - 48a_1^2a_2^2a_6 + 8a_1^2a_2a_4^2 - 30a_1^2a_3^2a_4 + 72a_1^2a_4a_6 + 16a_1a_2^2a_3a_4\\ &+ 36a_1a_2a_3^3 + 144a_1a_2a_3a_6 - 96a_1a_3a_4^2 - 16a_2^3a_3^2 - 64a_2^3a_6 + 16a_2^2a_4^2 + 72a_2a_3^2a_4\\ &+ 288a_2a_4a_6 - 27a_3^4 - 216a_3^2a_6 - 64a_4^3 - 432a_6^2,\end{aligned}$$

equipped with the point $\mathscr{O} = [0 : 1 : 0] \in E(K)$.

Note that this definition (1.9.2) is correct as stated in any characteristic for K, including $\operatorname{ch} K = 2, 3$. However, this leaves many questions unanswered. For instance, where does this monstrous formula for Δ come from, and how does one come up with it? More on this soon. The form of the curve in 1.9.3 is called the (projective) Weierstrass form, named after the 19$^{\text{th}}$ century German mathematician Karl Weierstrass (also spelled "Weierstraß", notably in German), who laid the foundation for the modern theory of elliptic curves.

Given the tools we have so far, the first definition (1.9.1) doesn't even make sense, since we haven't defined the genus of a curve (but see 1.8.17). I am also lying a little bit when I say that we defined everything else in the first definition, since I have only given you the definition of smooth projective *plane* curves; but there are curves more general than these, and we will meet some of these below (1.9.5(c)). Finally, I need to explain where the convention for naming the coefficients of the defining equation of an elliptic curve comes from; this I will explain below in 1.9.15(b). But before all of this, let me mention the

Fact 1.9.4. These two definitions of elliptic curves (1.9.1 and 1.9.2) are equivalent.

We will not need this fact, and so I will not prove it here. We will sketch below (1.9.9) the proof that every object of the kind mentioned in 1.9.2 satisfies the conditions of 1.9.1 (except for the genus one bit), but going the other direction really needs the Riemann-Roch Theorem; the interested reader can find the proof in [8, Proposition III.3.1]. We will adopt 1.9.2 as our official definition. This might raise the question of why bother with the first definition (1.9.1) at all; let me try to explain this now.

The basic idea is that smooth projective geometrically integral curves do show up in lots of other contexts. Let me give you a few examples (without proof).

Example 1.9.5.

(a) If $C \subset \mathbb{P}^2_K$ is a smooth curve and there is an $\mathscr{O} \in C(K)$, then $(C, \mathscr{O})$ is an elliptic curve according to 1.9.1. Indeed, smoothness and projectivity comes from the definition, geometric integrality come from a good salvage to 2.3.8(b), and the fact that C has genus one comes from, say, the degree-genus formula for smooth plane curves (see, e.g., [9, 18.6.7.1]). The transformation from C into Weierstrass form (1.9.3) is easy to achieve when $\mathscr{O} \in C(K)$ is an inflection point (1.9.6), but it harder to achieve when $\mathscr{O}$ is not an inflection point (you can find one way to do this, due to Nagell, in [1, §8]). This allows us to put lots of curves in Weierstrass form, and then use special techniques to study curves of this sort. This is exactly what happened when we were studying the (generalized) Fermat problem for $n = 3$ in §1.2.B; we found a change of coordinates that put $x^3+y^3 = \alpha$ (or more precisely its projective closure) into Weierstrass form to rephrase the problem. (We are yet to finish this proof.) We will also use 1.7.7 combined with our implication that 1.9.2 implies 1.9.1 to define the grofup law on an elliptic curve in Weierstrass form.

(b) Let K be a field. If a curve C is a smooth projective curve with a degree two cover of $\mathbb{P}^1_K$ branched at four points, then C has genus one, and is hence an elliptic curve in the sense of 1.9.1 if we can equip it with a K-rational point.[4] This is what enabled us to transform a curve of the form $v^2 = g(u)$ for $g(u) \in K[u]$ a quartic into Weierstrass form as in 2.1.12, giving us another way of approaching the Fermat problem for $n = 4$.

(c) Suppose we are interested in four-term sequences of integers $p < q < r < s$ such that p^2, q^2, r^2, s^2 are in arithmetic progression. (We already have enough tools to study such sequences of length three; see 2.3.4.) I hope that you agree that this is a reasonably natural question. To do this, we might look at the homogeneous equations

$$p^2 + r^2 = 2q^2 \text{ and } q^2 + s^2 = 2r^2.$$

These are homogeneous equations, and hence each cut out surfaces, say S_1 and S_2 in $\mathbb{P}^3_K$ when we equip it with homogeneous coordiantes $[p : q : r : s]$. It them turns out that the intersection $C := S_1 \cap S_2 \subset \mathbb{P}^3_{\mathbb{Q}}$ (at least in sufficiently large characteristic; I think ≥ 5 works) is a smooth projective geometrically integral curve of genus one ([9, 18.6.R]). Since it clearly has a rational point, say $[1 : -1 : -1 : 1] \in C(\mathbb{Q})$, this tells us that the curve C is an elliptic curve in the sense of 1.9.1, and so the abstract theory guarantees that we can put it in Weierstrass form 1.9.2. If we can make this transformation explicit, then this give us a way to understand four-term sequences of squares in arithmetic progression. I invite you to do this in an elementary fashion in 2.3.5.

One general cultural remark (really due to Dan Shapiro): if you're studying a new field of mathematics, then you should think about questions or theorems, the statements of which do not involve any concepts from your theory, but the "most natural" answers or proofs of which use the theory in an essential way. Further, if you are having a really hard time coming up with examples of such things, then you should carefully reevaluate why you might be interested in studying that particular theory. I hope that the Fermat examples and the example of 1.9.5(c) can provide a few such natural questions which might motivate you to study the theory of elliptic curves.

In lieu of proving 1.9.4 in full generality, let me at least tell you how to transform a smooth cubic $C \subset \mathbb{P}^2_K$ over some field K with an inflection point $\mathscr{O} \in C(K)$ into Weierstrass form by a change of coordinates.

Lemma 1.9.6. Let K be a field, $C \subset \mathbb{P}^2_K$ be a smooth cubic, and let $\mathscr{O} \in C(K)$ be an *inflection point*. Then there is a projective change of coordinates that brings C into the projective Weierstrass form (1.9.2), with $\mathscr{O}$ mapping to $[0 : 1 : 0]$.

Recall that $\mathscr{O} \in C(K)$ is said to be an *inflection point* or a *flex point* iff $\mathbb{T}_{\mathscr{O}}C$ intersects C with multiplicity 3 at $\mathscr{O}$, or equivalently iff $\mathscr{O} * \mathscr{O} = \mathscr{O}$. The following proof is very similar to that of 1.6.4, which might be helpful to review now.

[4] To be more precise, I should say: the morphism $C \to \mathbb{P}^1_K$ should be finite separable of degree two, where the ramification locus $R \subset \mathbb{P}^1_K$ is a closed subscheme of length 4 with at least one K-rational point in $R(K)$. If you have no idea what this means, you can safely ignore this remark.

Proof. Take a projective change of coordinates which sends $\mathscr{O}$ to $[0:1:0]$, and sends $\mathbb{T}_{\mathscr{O}}C$ to the line $L_\infty = \mathbb{V}(Z)$ at infinity. In this coordinates system, pick a defining equation F for C, and express it in the form

$$F = A_0Y^3 + A_1Y^2 + A_2Y + A_3,$$

where the $A_i \in K[X,Z]$ is homogeneous of degree i for $i = 0,\dots,3$. The fact that $\mathscr{O} = [0:1:0] \in C(K)$ implies that $A_0 = 0$. The fact that $\mathbb{T}_{\mathscr{O}}C = \mathbb{V}(Z)$ implies (check!) that A_1 is a nonzero scalar multiple of Z. Therefore, by rescaling F appropriately, we may assume that $A_1 = Z$. Finally, the fact that $\mathscr{O} = [0:1:0]$ is an inflection point on C implies that the tangent line $L_\infty = \mathbb{V}(Z)$ intersects C with multiplicity three at $\mathscr{O}$, which forces that $Z \mid A_2$ (check!). This means that F has the form

$$Y^2Z + (a_1X + a_3Z)YZ - (a_0X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3)$$

for some $a_0, a_1, a_2, a_3, a_4, a_6 \in K$. Can a_0 be zero? If this is the case, then $Z \mid F$, and so C is not integral; but this can't happen if C is smooth, basically by Bézout's Theorem (see 1.7.1 and 2.3.8(b)). Therefore, $a_0 \neq 0$, and we may multiply F throughout by a_0^2 and let $(X', Y') := (a_0X, a_0Y)$ to get an equation in Weierstrass form (1.9.2), up to minor relabeling of the coefficients. It remains to justify that the smoothness of C implies that $\Delta \neq 0$; this we do in a more general context below (1.9.9). ∎

Exercise 1.9.7. Fill in the gaps in the above proof: check that $A_0 = 0$ follows from $\mathscr{O} \in C(K)$, check that $A_1 = Z$ (up to scaling) follows from $\mathbb{T}_{\mathscr{O}}C = \mathbb{V}(Z)$, and check that $Z \mid A_2$ follows from the fact that $\mathscr{O} \in C(K)$ is an inflection point. It may be helpful to use local coordinates $(t,s) = (X/Y, Z/Y)$ around the point $(0,0)$.

Exercise 1.9.8. Let K be a field of characteristic other than 3, and let $\alpha \in K^\times$. Show that the curve $E_\alpha\mathbb{V}(X^3 + Y^3 - \alpha Z^3) \subset \mathbb{P}^2_K$ of §1.2.B and 2.2.4 has an inflection point at $[1:-1:0] \in E_\alpha(K)$, and that the linear change of coordinates 1.2.8 is obtained exactly by the procedure described in the proof of 1.9.6, up to minor modifications (which you should figure out).

Let us now sketch the proof of the result I promised which implies that a curve in Weierstrass form (1.9.2) has the properties of 1.9.1, except for the part involving the genus, which we will not need (and a proof of which can be found at [9, 18.6.7.1]).

Proposition 1.9.9. Let K be a field (of any characteristic!) and $E \subset \mathbb{P}^2_K$ be a curve in Weierstrass form, as described in 1.9.2. Then E is smooth iff $\Delta \neq 0$. In particular, if this is the case, then E is a smooth projective geometrically integral curve of genus one, with $\mathscr{O} = [0:1:0] \in E(K)$, i.e., an elliptic curve in the sense of 1.9.1.

A few remarks are in order. Firstly, as I mentioned above, I will not prove the part involving the genus. Secondly, as before, I will invoke 2.3.8(b) to reduce to showing only that E is smooth iff $\Delta \neq 0$. I don't feel too bad about this, because we might later show that any curve in Weierstrass form (smooth or not!) is automatically geometrically integral. Thirdly, we will also show later a simpler version of this statement in characteristic other than 2 or 3. Finally, I will only give the key ideas behind the proof of 1.9.9, and leave some details to the reader. A full proof, using slightly different terminology, can be found at [8, Proposition III.1.4(a)].

Proof Sketch of 1.9.9. Let's first begin by study the points at infinity. Setting $Z = 0$ in 1.9.3 gives us $X^3 = 0$, which shows that $\mathscr{O} = [0:1:0] \in E(K)$ is the unique point at infinity on a curve in Weierstrass form. Either by using the coordinates $(t,s) = (X/Y, Z/Y)$, or by taking $\partial F/\partial Z$, it is easy to see (check!) that $\mathscr{O}$ is a smooth point with tangent line $L_\infty = \mathbb{V}(Z)$. (In fact, the fact that setting $Z = 0$ gives us $X^3 = 0$ tells us that $\mathscr{O}$ is an inflection point on $E(K)$, but we won't need this.)

Having checked the unique point at infinity for smoothness (which is true irrespective of whether $\Delta \neq 0$), we can now restrict to studying the affine equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{1.9.10}$$

which corresponds to the vanishing of $f := y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6 \in K[x,y]$. To check smoothness, we evaluate the partial derivatives

$$\frac{\partial f}{\partial x} = a_1y - 3x^2 - 2a_2x - a_4 \text{ and } \frac{\partial f}{\partial y} = 2y + a_1x + a_3. \tag{1.9.11}$$

For simplicity, first assume that $\operatorname{ch} K \neq 2, 3$ and that $a_1 \neq 0$. Then if $\partial f/\partial y = 0$, the we can express y in terms of x as

$$y = -\left(\frac{a_1 x + a_3}{2}\right). \tag{1.9.12}$$

Plugging this into the expression for $\partial f/\partial x$ in 1.9.11 yields a quadratic equation for x, which we can solve for using the quadratic formula. Having solved for x, we can then use 1.9.12 to express y in terms x and plug both into 1.9.10 to get a huge equation involving just one huge square root. We can then move this square root to one side, take squares, and rearrange along with some very minor modifications to arrive precisely at Δ. This tells us that there is a singular point iff $\Delta = 0$. (This procedure might explain the huge size of Δ.)

This proof doesn't quite cover all cases. We still need to worry about small characteristics and whether $a_1 = 0$. For the purposes of illustration, let me do one more case: that of $\operatorname{ch} K = 2$ and $a_1 = 0$. Note that in characteristic two, the expression for Δ simplifies tremendously to

$$\Delta = a_1^6 a_6 + a_1^5 a_3 a_4 + a_1^4 a_2 a_3^2 + a_1^4 a_4^2 + a_1^3 a_3^3 + a_3^4. \tag{1.9.13}$$

In particular, if further $a_1 = 0$, then $\Delta = a_3^4$ simply. If there is a singular point, then $\partial f/\partial y = 0$ implies $a_3 = 0$ and hence $\Delta = 0$. Conversely, if $\Delta = 0$, then $\partial f/\partial y = 0$ identically. The expression for $\partial f/\partial x$ simplifies to $x^2 + a_4$, so that we may solve to get $x = a_4^{1/2}$. Finally, plugging $x = a_4^{1/2}$ into 1.9.10 yields the quadratic equation (check!)

$$y^2 + (a_1 a_4^{1/2} + a_3)y + a_2 a_4 + a_6 = 0,$$

which can certainly be solved for y (over an algebraically closed field $\Omega \supset K$!), yielding a singular point $(x, y) \in E(\Omega)$, and showing that E is not smooth.

The other cases are similar and left to the reader as an extended exercise. ■

Exercise 1.9.14. Do a few more cases to convince yourself that the result is true as claimed in all cases. Then, if you feel the need to, look up the general proof in [8, Proposition III.1.4(a)].

Remark 1.9.15.

(a) The affine equation 1.9.10 is what is often called the *Weierstrass form* of the elliptic curve. When we say "an elliptic curve with Weierstrass equation 1.9.10", what is really meant is the projective closure 1.9.3, which is an elliptic curve iff $\Delta \neq 0$. In particular, even though we'll work with the affine equation 1.9.10, it should be kept in mind that the curve $E \subset \mathbb{P}^2_K$ we are considering is a projective curve with the point $\mathscr{O} = [0 : 1 : 0] \in E(K)$ at infinity.

(b) The affine equation 1.9.10 explains the notation for the coefficients a_i: if we assign the weight 2 to x and 3 to y, then 1.9.10 becomes a homogeneous equation iff we assing weight i to a_i for $i = 1, 2, 3, 4, 6$. This also explains why there is no a_5. There are many reasons why we might want to do this; for now, it suffices to note that this will make every constant we define come with a natural weight. For instance, you can check that although Δ (given in 1.9.2) is not homogeneous in $a_1, \ldots, a_6$ if we assign the weight 1 to each a_i, it *is* homogeneous of degree 12 if we assign weight i to a_i. This might be helpful to keep in mind when we define the constants c_4 and j later.

Unimportant Remark 1.9.16. The above discussion may not fully convincing about the real origins of Δ, even though it is easier to check that this statement (1.9.9) is correct. Indeed, the formula for Δ in 1.9.2 really comes from *elimination theory*, which answers the following question: what is the "best" way to eliminate the variables x and y from the system of equations (1.9.10) and (1.9.11)? This amounts to finding generators for the ideal J given as the intersection

$$J := I \cap \mathbb{Z}[a_1, a_2, a_3, a_4, a_6] \subset \mathbb{Z}[a_1, a_2, a_3, a_4, a_6], \tag{1.9.17}$$

where I is the ideal

$$I = \left(f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \in \mathbb{Z}[a_1, \ldots, a_6, x, y].$$

The ideal J of 1.9.17 is known as the *elimination ideal*, and it can be shown, say using the theory of Gröbner bases, that J is principal, generated precisely by Δ. (If you don't trust me–and indeed, you shouldn't be willing to take things on faith–then you can easily check this fact in software such as, say,

Macaulay2. There is a good geometric reason why the ideal J should be principal; I'm not going to include this here, but if you are interested, you should ask me.) This is where Δ really comes from–it is the generator of the elimination ideal. If you want to read more about elimination theory, you can find it in [10] when working over fields and [11, Chapter 4] over more general rings such as $\mathbb{Z}$ (which is basically what Macaulay2 is actually doing).

In closing today, I wish to remark that when the characteristic of K is not small, then we may come up with changes of coordinates to simply 1.9.10 even further. For instance, if $\operatorname{ch} K \neq 2$, then we may complete the square on the left hand side of 1.9.10 by taking $x_1 = x$ and

$$y_1 = y + \frac{a_1 x + a_3}{2}$$

to get a shorter Weierstrass equation of the form

$$y_1^2 = x_1^3 + a x_1^2 + b x_1 + c. \tag{1.9.18}$$

Note that you *cannot* do something along these lines $\operatorname{ch} K = 2$, since the curve 1.9.18 is necessarily singular in characteristic two–check!. Finally, if $\operatorname{ch} K \neq 2, 3$, then we may simplify 1.9.18 even further by taking $(x_2, y_2) = (x_1 + (a/3), y_1)$ to depress the cubic, yielding an even shorter Weierstrass equation of the form

$$y_2^2 = x_2^3 + b' x_2 + c'. \tag{1.9.19}$$

Again, you may not be able to do this if $\operatorname{ch} K = 3$. We will study these shorter Weierstrass forms in more detail next time.

Remark 1.9.20. One can check that Δ remains invariant under the operations of the form that allowed to us bring curves into short Weierstrass form. This is almost miraculous. For an example of what I mean by this: if we use that substitution $(x_1, y_1) = (x, y + (a_1 x + a_3)/2)$ when $\operatorname{ch} K \neq 2$, then you can check that 1.9.10 transforms into 1.9.18, where

$$(a, b, c) = \left(a_2 + \frac{1}{4} a_1^2, a_4 + \frac{1}{2} a_1 a_3, a_6 + \frac{1}{4} a_3^2 \right).$$

(Check this!) What I mean by the invariance of Δ is then that

$$\Delta(a_1, a_2, a_3, a_4, a_6) = \Delta(0, a, 0, b, c) := \Delta\left(0, a_2 + \frac{1}{4} a_1^2, 0, a_4 + \frac{1}{2} a_1 a_3, a_6 + \frac{1}{4} a_3^2\right),$$

both sides being equal to

$$\Delta = -16(4a^3 c - a^2 b^2 - 18abc + 4b^3 + 27c^2). \tag{1.9.21}$$

A similar thing happens with the substitution that depressed the cubic to be of the form 1.9.19. (Check this!) This is something special to the mysterious polynomial Δ, and obviously not true of all polynomials in the a_i. This mysterious invariance property of Δ is incredibly helpful when working with shorter Weierstrass forms 1.9.18 or 1.9.19, because then we can just use the simplified expression 1.9.21.

Unimportant Remark 1.9.22.

(a) For analogs of these shorter Weierstrass forms when $\operatorname{ch} K$ is 2 or 3, you may consult [8, Appendix A] or [12, Chapter 3].

(b) One might ask why I have bothered at all to define the longer Weierstrass form 1.9.10 at all, when we could work with the shorter forms 1.9.18 or 1.9.19, as is done, for instance, in [2]. There are a few reasons for this. One of them is that the longest Weierstrass form 1.9.10 works in *any* characteristic; in particular, it works in characteristics 2 and 3, in which the shorter ones might not. Secondly, software such as Sage is used to taking input into the form of a five-tuple $(a_1, a_2, a_3, a_4, a_6)$ produce elliptic curves, and I wanted you to understand what Sage is doing. Finally, even when working exclusively with elliptic curves over $\mathbb{Q}$, we sometimes need the long Weierstrass form 1.9.10, for instance, when finding convenient forms in which to express 2-isogenies (more on this soon!), or trying to find minimal models to study primes of bad reduction, or when giving short forms in which to express the equation of the curve when the coefficients become huge, as is needed for the curves involved in state-of-the-art records on elliptic curve ranks over $\mathbb{Q}$ (more on this later!). Even though it might be easiest to start with the shortest Weierstrass form, I think that it is not much more difficult pedagogically to start with the longer Weierstrass form 1.9.10, which allows an easier transition into topics mentioned in this remark.