1.17 07/17/24 - Classification of Elliptic Curves, Story Time

Recall our standing assumption that the base field k is algebraically closed of characteristic other than 2 or 3. We showed in lecture last time that every elliptic curve (E, O) with $O \in E$ an inflectionary point can be put into Legendre form, i.e. there is some $\lambda \in k \smallsetminus \{0, 1\}$ and a projective change of coordinates taking E to the curve E_λ which is the projective closure of the affine curve defined by

$$y^2 = x(x-1)(x-\lambda)$$

with basepoint $O = [0 : 1 : 0]$. For a given E, how many such λ work? In other words, when are the curves E_λ and E_μ for $\lambda, \mu \in k \smallsetminus \{0, 1\}$ isomorphic (i.e. related by a change of coordinates)? Answering this question will enable us to "classify" all elliptic curves in the sense that we will be able to tell exactly when two such cubics are related by a change of coordinates, somewhat similarly to Theorem 1.12.13. For this, we need to introduce a key invariant of elliptic curves–the j-invariant.

1.17.1 The j-Invariant

I have used the word "isomorphism" quite a few times already; let me explain what I mean by that at the moment.

Definition 1.17.1.

(a) Two curves $D, E \subset \mathbb{P}^2_k$ are said to be projectively isomorphic if there is a projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ such that $\Phi(D) = E$.

(b) Let (E, O) and (E', O') be two plane elliptic curves. We say that (E, O) and (E', O') are weakly isomorphic if the underlying cubic curves $E, E' \subset \mathbb{P}^2_k$ are projectively isomorphic; we say that (E, O) and (E', O') are strongly isomorphic if the projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ taking $\Phi(E) = E'$ can be chosen to satisfy $\Phi(O) = O'$. Such a Φ is called a strong isomorphism $(E, O) \to (E', O')$.

Note that projective isomorphism (i.e. being projectively isomorphic) is an equivalence condition (often denoted by $\cong$) on the set of all projective plane curves; projectively isomorphic curves have the same number and degrees of irreducible components, singular points, etc. We will often denote projective isomorphism via the notation $\cong$.Effectively, they are "the same" curve, just viewed under different coordinates. Note also that strongly isomorphic elliptic curves are weakly isomorphic, and a strong isomorphism Φ is automatically a group isomorphism thanks to the geometric nature of the group law on E. Finally, if $E \subset \mathbb{P}^2_k$ is a smooth cubic, and $O \in E$ an inflection point while $O' \in E$ *not* an inflection point, then the elliptic curves (E, O) and (E, O') are *not* strongly isomorphic, since the property of being an inflection point of a curve is preserved under projective changes of coordinates.

Remark 1.17.2. The terminology here is my own and not standard. Further, in slightly more advanced treatments of algebraic geometry, there is yet another notion of isomorphism: the notion of an "abstract" isomorphism of algebraic curves, given by polynomial or rational functions. For instance, the parametrization in Example 1.13.3 combined with the projection explained in Lecture 1.3 tells us that a line $L \subset \mathbb{P}^2_k$ and a smooth conic $C \subset \mathbb{P}^2_k$ are abstractly isomorphic, although they cannot be projectively isomorphic because they have different degrees. However, it turns out that for elliptic curves the notions of abstractly isomorphic and projectively isomorphic agree, although showing this needs more work. (See the grown-up text [9] if you are curious.) Here we will restrict ourselves to the study of projective isomorphisms.

Finally, one last definition that we will need is

Definition 1.17.3. The j-function $j : k \smallsetminus \{0,1\} \to k$ is the rational function defined by

$$j(\lambda) := 2^8 \frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(\lambda - 1)^2}.$$

The origin of this mysterious function will be explained in Remark 1.17.7 below; before that, we arrive at the (somewhat surprising) main result of this section.

Theorem 1.17.4. Let $\lambda, \mu \in k \smallsetminus \{0,1\}$, and let $O := [0:1:0]$ be the usual basepoint. The following are equivalent:

(a) The elliptic curves (E_λ, O) and (E_μ, O) are strongly isomorphic.
(b) The curves $E_\lambda, E_\mu \subset \mathbb{P}^2_k$ are projectively isomorphic.
(c) We have $\mu \in M_\lambda := \{\lambda, 1/\lambda, 1-\lambda, 1/(1-\lambda), \lambda/(\lambda-1), (\lambda-1)/\lambda\}$.
(d) We have $j(\lambda) = j(\mu)$.

Proof. For this, we need the following two observations:

(1) Let $E \subset \mathbb{P}^2_k$ be a smooth cubic curve, and $O, O' \in E$ be two inflection points. Then the elliptic curves (E, O) and (E, O') are strongly isomorphic. For a proof, consider the line $L_{O,O'}$, and let the third point of its intersection with E be O''; then $O'' \in E$ is also an inflection point by Exercise 2.5.5 (b), and $O'' \neq O, O'$ if $O \neq O'$. Put (E, O'') in Weierstrass normal form; then since O and O' are collinear with O'', we have on this elliptic curve that $O + O' = 0$. It follows that the projective change of coordinates $Y \mapsto -Y$ is the required strong isomorphism between (E, O) and (E, O').
(2) If $E \subset \mathbb{P}^2_k$ is a smooth cubic curve and $O \in E$ an inflection point, then except for the flex tangent there are three other tangents to E that pass through O, and the points of contact of these three lines with E are collinear. This is immediate by putting (E, O) in Weierstrass or Legendre form.

We are now ready to proceed to the main proof.

(a) $\Leftrightarrow$ (b) Strongly isomorphic elliptic curves are weakly isomorphic by definition. follows from (1). Indeed, if $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ is a change of coordinates such that $\Phi(E_\lambda) = E_\mu$, then $\Phi(O) \in E_\mu$ is an inflection point. By (1), there is a strong isomorphism Ψ taking $(E_\mu, \Phi(O))$ to (E_μ, O). Then $\Psi \circ \Phi$ is a strong isomorphism taking (E_λ, O) and (E_μ, O).

(a) $\Leftrightarrow$ (c) For one direction, let Φ be the strong isomorphism taking (E_λ, O) to (E_μ, O). Then Φ takes $\mathrm{T}_O E_\lambda$ to $\mathrm{T}_O E_\mu$, i.e. preserves the line $Z = 0$ as a set (although not necessarily pointwise). By (2), Φ must take the set the points $\{(0,0), (1,0), (\lambda, 0)\}$ to $\{(0,0), (1,0), (\mu, 0)\}$. In particular, Φ must fix the line $Y = 0$ as well, and hence the point $[1:0:0]$. This combined with the fact that $\Phi(O) = O$ implies that Φ must be of the form $[X:Y:Z] \mapsto [sX + tZ : Y : Z]$ for some $s, t \in k$ with $s \neq 0$ (check!). In particular, the automorphism $x \mapsto sx + t$ takes the set $\{0, 1, \lambda\}$ to $\{0, 1, \mu\}$. In particular, $t \in \{0, 1, \mu\}$, and for each choice of t, we are left with two possibilities for s; these correspond to the six choices for μ above. Conversely, the same argument shows that when $\mu \in M_\lambda$, a transformation of this sort gives us the required strong isomorphism.

(c) $\Leftrightarrow$ (d) This follows from the identity

$$j(\lambda) - j(\mu) = 2^8 \frac{(\lambda - \mu)(\lambda\mu - 1)(\lambda + \mu - 1)(\lambda\mu - \mu + 1)(\lambda\mu - \lambda + 1)(\lambda\mu - \lambda - \mu)}{\lambda^2(\lambda - 1)^2\mu^2(\mu - 1)^2}.$$

■

This key theorem allows us to define a crucial invariant for smooth cubic curves: the j-invariant.

Definition 1.17.5. Let $E \subset \mathbb{P}^2_k$ be a smooth cubic curve. Define the j-invariant of E as follows: pick a projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ such that $\Phi(E)$ is in Legendre form, say $\Phi(E) = E_\lambda$, and define

$$j(E) := j(\lambda).$$

That this is well-defined follows from Theorem 1.17.4. From this we finally arrive at the required classification theorem for all cubic curves.

Corollary 1.17.6. Up to projective changes of coordinates, a cubic curve in $\mathbb{P}^2_k$ is of exactly one of the following seven types.

(a) The union of three concurrent lines.
(b) The union of three nonconcurrent lines.
(c) The union of a smooth conic and a line tangent to it.
(d) The union of a smooth conic and a line transverse to it (i.e. meeting it in two distinct points).
(e) A nodal cubic curve.
(f) A cuspidal cubic curve.
(g) A smooth cubic curve, i.e. after choosing a basepoint, an elliptic curve.

Further:

(a) The types (a) - (d) correspond to the reducible cubics, and the types (e) - (g) to the irreducible cubics. Of these, all curves of types (a) - (f) are singular.
(b) Any two curves of the same type from (a) - (f) are projectively isomorphic.
(c) Two smooth cubic curves are projectively isomorphic iff they have the same j-invariant. Further, given any specified $\alpha \in k$, there is a smooth cubic $E \subset \mathbb{P}^2_k$ with j-invariant α.

In particular, the j-invariant is a complete isomorphism invariant of smooth cubic curves, and can take any value in k.

Proof. The case of the reducible cubics is easy and left to the reader and the case of the irreducible but singular cubics was handled in Exercise 2.4.4, so we'll take the classification as well as statements (a) and (b) as proven. For (c), it is firstly clear that the j-invariant of smooth cubic is a projective isomorphism invariant; this is the content of Theorem 1.17.4. Conversely, suppose that $E, E' \subset \mathbb{P}^2_k$ are two smooth cubic curves with $j(E) = j(E')$. By the discussion in §1.16.1, there are $\lambda, \mu \in k \smallsetminus \{0, 1\}$ and projective isomorphisms $E \cong E_\lambda$ and $E' \cong E_\mu$. It follows then from the first part that

$$j(\lambda) = j(E_\lambda) = j(E) = j(E') = j(E_\mu) = j(\mu),$$

so from Theorem 1.17.4 we conclude that $E_\lambda \cong E_\mu$. It then follows that

$$E \cong E_\lambda \cong E_\mu \cong E'$$

as needed. Finally, given an $\alpha \in k$, solve the equation

$$2^8(\lambda^2 - \lambda + 1)^8 - \alpha\lambda^2(\lambda - 1)^2 = 0$$

to get a $\lambda \in k \smallsetminus \{0, 1\}$ (using $k = \overline{k}$ and $\operatorname{ch} k \neq 2$); then $E_\lambda \subset \mathbb{P}^2_k$ is a smooth cubic with j-invariant α. ∎

Remark 1.17.7. In a more advanced perspective on the theory of elliptic curves, it is seen that elliptic curves are $2:1$ covers of $\mathbb{P}^1_k$ branched over four points, and the location of the 4 points in $\mathbb{P}^1_k$ (up to projective changes) determines the isomorphism type of corresponding elliptic curve. In the above set-up (i.e. when E is in say Legendre form $y^2 = x(x-1)(x-\lambda)$), this map to $\mathbb{P}^1_k$ is given by taking the x-coordinate; for most values of x, there are two values of y, i.e. two points in E, mapping to it–except for the values $x = 0, 1, \infty, \lambda$. Now given an ordered quadruple of points (a, b, c, d) of $\mathbb{P}^1_k$, we can associate to them a quantity–the cross ratio–which is invariant under coordinate changes; however, permuting the 4 points gives rise to up to six different numbers, each of which is an equal candidate for the title of the cross ratio of an unordered quadruple of points. To systematize this, we can note that any four points on $\mathbb{P}^1_k$ can be brought via a projective change of coordinates into a tuple of the form $(0, 1, \infty, \lambda)$ for some $\lambda \in k \smallsetminus \{0, 1\} = \mathbb{P}^1_k \smallsetminus \{0, 1, \infty\}$–and indeed, this λ then *is* the cross-ratio. The set M_λ is the set of values $\mu \in \mathbb{P}^1_k \smallsetminus \{0, 1, \infty\}$ such that the quadruple $\{0, 1, \infty, \mu\}$ has the same cross-ratio as that of $\{0, 1, \infty, \lambda\}$ when taken in some order, and as the proof of Theorem 1.17.4 shows, the j-function captures precisely this set, and provides a true invariant (under coordinate changes) of unordered quadruples of points on $\mathbb{P}^1_k$. In more grown-up terminology, there is an S_3 action on $\mathbb{P}^1_k$, the orbit of a fixed $\lambda \in \mathbb{P}^1_k \smallsetminus \{0, 1, \infty\}$ under this action is exactly M_λ, and the j function $j : \mathbb{P}^1_k \to \mathbb{P}^1_k$ is a rational function of degree 6 that exhibits $\mathbb{P}^1_k$ as the quotient $\mathbb{P}^1_k / S_3$. (The factor of 2^8 is there for further normalization purposes; I will not explain here what that means.)

Remark 1.17.8. The most interesting values of λ are the ones for which the set M_λ has fewer than 6 elements; these correspond to elliptic curves E_λ with additional symmetries. A little computation shows that there are (in $\operatorname{ch} k \neq 2, 3$) exactly 5 such values corresponding to

$$M_{-1} = M_2 = M_{1/2} = \{-1, 1/2, 2\} \text{ and } M_\rho = M_{\overline{\rho}} = \{\rho, \overline{\rho}\},$$

where ρ is a primitive sixth root of unity, i.e. $\rho^2 - \rho + 1 = 0$, and $\overline{\rho} = 1 - \rho$. In the former case, we are looking at the elliptic curve $y^2 = x^3 - x$ which has $j = 1728$ and has a $\mathbb{Z}/4$-symmetry $(x, y) \mapsto (-x, \mathrm{i}y)$; in the latter case, we are looking at the curve $y^2 = x^3 - 1$ which has $j = 0$ and the $\mathbb{Z}/6$-symmetry $(x, y) \mapsto (\rho^2 x, \rho^3 y)$. This is (part of) the reason for the specialness of the values $j = 0$ and $j = 1728$ in the theory of elliptic curves. (Surprisingly, for $\operatorname{ch} k = 2, 3$ curves with $j = 0 = 1728$, we have automorphism groups $\mathrm{SL}_2 \mathbb{F}_3$ and $\mathbb{Z}/3 \rtimes \mathbb{Z}/4$ of sizes 24 and 12 respectively; see [9, Theorem 10.1].)

There is, of course, much more to say about elliptic curves, but that is all we have time for, because I want to spend some time narrating some stories.

1.17.2 Story Time

The theory of plane algebraic curves is a classical yet very rich subject which is both the starting point of several deep stories (the theory of abstract curves, algebraic geometry in general, and elliptic curves to name a few) and the source of still many unsolved problems and not-as-well-understood phenomenen (e.g. the moduli of curves). In this section, I want to end the course by mentioning some of the directions the study of curves can take from here.

- For starters, there is much, much more to say about elliptic curves. For a good introduction you can start reading now, see [10]; for a more advanced perspective, the classical textbook is [9]. Here are three facts I want to mention:
  (a) Over the field $k = \mathbb{C}$, an elliptic curve $E \subset \mathbb{P}^2_{\mathbb{C}}$ is "the same" as a complex torus, i.e. $\mathbb{C}/\Lambda$ for some lattice $\Lambda \subset \mathbb{C}$, relating the theory of plane cubics to doubly periodic meromorphic functions in the plane (which is–via the theory of elliptic integrals–ultimately the source of the nomenclature "elliptic", since it is otherwise not so clear what these curves have to do with ellipses). The addition law on E is then

induced from the usual addition law on $\mathbb{C}$ (or more precisely on the quotient group $\mathbb{C}/\Lambda$); this perspective makes abundantly clear why for each $n \geq 1$ we can expect $E[n] \cong \mathbb{Z}/n \times \mathbb{Z}/n$–these correspond exactly to the n^2 points in the fundamental parallelogram of Λ corresponding to $(1/n)\Lambda$. Further, if for each τ in the upper half plane $\mathbb{H}$, we let E_τ be the elliptic curve corresponding to the lattice $\mathbb{Z} \oplus \mathbb{Z}\tau$, then the function $\tau \mapsto j(E_\tau)$ is a holomorphic function $\mathbb{H} \to \mathbb{C}$ which very beautiful properties (it is invariant under the action of the modular group $\mathrm{PSL}_2\,\mathbb{Z}$, and intimately related to the theory of modular forms etc.).

(b) Elliptic curves over finite fields $k = \mathbb{F}_q$ are both theoretically important and the key to a lot of modern day cryptography. For instance, the Hasse bound says that if $E \subset \mathbb{P}^2$ is an elliptic curve defined over $\mathbb{F}_q$, then the number $\#E(\mathbb{F}_q)$ of $\mathbb{F}_q$-points on E is bound by

$$|\#E(\mathbb{F}_q) - (q+1)| \leq 2\sqrt{q}.$$

Such counts of points an elliptic curve for varying $q = p^n$ are the start of another beautiful story–that of the Weil conjectures.

(c) Finally, over number fields such as $k = \mathbb{Q}$ the theory is still fascinating and at times mysterious. The famous Mordell-Weil Theorem asserts that if k is a number field and E an elliptic curve defined over k, then the group of k-rational points $E(k)$ is a finitely generated abelian group. In particular, it is isomorphic to $\mathbb{Z}^r \oplus T$ for some unique integer $r \geq 0$ and finite abelian group T (the torsion subgroup)–this r is called the algebraic rank of $E(k)$. A theorem of Barry Mazur (a professor of mine!) asserts that when $k = \mathbb{Q}$, the torsion subgroup T can be only one of 15 types–it can be either $\mathbb{Z}/n$ for $n = 1, \ldots, 10$ or $n = 12$, or it can be $\mathbb{Z}/2n \times \mathbb{Z}/2$ for $n = 1, 2, 3, 4$. Much work has been done to extend this result to general number fields k. Finally, the rank r is another fascinating quantity. The largest known rank over $k = \mathbb{Q}$ as of this writing (September 2024) is 29, and this elliptic curve was found by Elkies and Klagsbrun (Elkies was another professor of mine); see [11] for the equation. The rank is part of some very important unsolved conjectures as well: associated to an elliptic curve E over $k = \mathbb{Q}$, we also have a an L-function holomorphic in a neighborhood of $s = 1$; the order of vanishing of this function at $s = 1$ is called the analytic rank of E. The famous Birch and Swinnerton-Dyer conjecture asserts that the analytic rank of an elliptic curve agrees with its algebraic rank–if you show this, you get a million dollars (among other things)!

- Another theme that is still an area of active research that touched upon in Exercise 2.1.2 is real algebraic geometry. This field studies the topology and geometry of algebraic curves (and, more generally, varieties) over the field $k = \mathbb{R}$ of real numbers, which is harder than the case $k = \mathbb{C}$ because $\mathbb{R}$ is not algebraically closed. It is theorem due to Harnack from the 19th century that a real projective algebraic curve of degree $d \geq 1$ has at most $\binom{d-1}{2}+1$ connected components; for a proof sketch, see [12, Lect. 19]. Harnack also showed with this proof that this bound is actually achieved–curves with this maximal number of connected components are called M-curves. The classification of all possible isotopy types (roughly, the nesting type) of M-curves still remains quite mysterious for $d \geq 8$. In his list of 23 mathematical problems presented by Hilbert before the Paris conference of the International Congress of Mathematicians in 1900, the study of real algebraic geometry was in the sixteenth place, and this problem has occupied researchers fruitfully for almost a century with lots still to be explored.
- In our focus on smooth curves, one fascinating area of study we completely missed out on was singularity theory, which studies the singularities of curves (or more generally varieties). For instance, if you look at the nodal curve C defined by $y^2 = x^3 - x^2$ over $k = \mathbb{C}$ and take a small cross-section near the singularity $(0,0)$ in $\mathbb{A}^2_{\mathbb{C}} \cong \mathbb{C}^2$ by a 3-sphere $S^3 \subset \mathbb{C}^2$, the intersection $C \cap S^3$ is a link in S^3, namely the Hopf link–two circles that don't intersect

but cannot be "pulled apart" because they link exactly once. Similarly, if C were to be the cuspidal cubic $y^2 = x^3$, then the intersection $C \cap S^3$ would be the trefoil knot. Therefore, the resulting link $C \cap S^3$ in S^3 is somehow capturing the nature of the singularity of the corresponding curve C–the more complicated the singularity, the more complicated the corresponding link (this observation was first made by K. Brauner). Knots and links arising in this way are called knots of singularities, and were studied extensively by Milnor, the standard resource on the subject still being [13]. Here's one thing to think about: the figure-8 knot 4_1 does *not* arise as a complex knot singularity. Can you think of how you would prove something like this?

- One more connection I want to mention, already somewhat manifest in our discussion of elliptic curves over $k = \mathbb{C}$ above, is the relationship between algebraic geometry and complex analysis. Classically, these subjects were not considered separate at all, with the main focus being the study of complex algebraic curves. The idea is that if $X \subset \mathbb{P}^2_{\mathbb{C}}$ is a smooth curve, then X is a compact complex manifold of dimension one–a Riemann surface. By a topological classification theorem of orientable closed surfaces, each such X is a g-holed torus for some $g \geq 0$ (think of a the surface of a donut with g holes; the case $g = 0$ is the sphere, and $g = 1$ the standard torus). This integer g–called the genus of the curve X–is directly computable from X and contains a lot of information about it. If $X \subset \mathbb{P}^2_{\mathbb{C}}$ is a smooth curve of degree d, then the genus of X can be shown to be $\binom{d-1}{2}$. One piece of information contained in g is about a natural geometry on X. It turns out that any compact Riemann surface carries in a natural way a metric, which for $g = 0, 1, \geq 2$ corresponds to round, flat, and hyperbolic metrics. The round case corresponds to $g = 0$ being $X \cong \mathbb{P}^1_{\mathbb{C}}$, which is topologically a sphere (the Reimann sphere). The flat case $g = 1$ corresponds to the case of plane cubics–$d = 3$–which, as mentioned above, are naturally isomorphic to Riemann surfaces of the form $\mathbb{C}/\Lambda$ for lattices $\Lambda \subset \mathbb{C}$; the flat metric on the elliptic curve then comes from the Λ-translation-invariant flat metric on $\mathbb{C}$. By far the most interesting and mysterious case is $g \geq 2$, when we have a natural hyperbolic metric on X, i.e. a metric of constant negative curvature. Much work has been done to study the moduli theory of such curves, although a complete understanding is far beyond our means at the moment. Finally, a famous theorem (a generalization of which is due to Chow) asserts conversely that any complex submanifold $X \subset \mathbb{P}^2_{\mathbb{C}}$ is a smooth algebraic curve, so in dimension 1 the theory of compact complex manifolds and smooth algebraic varieties (i.e. curves) are identical (although these theories, importantly, diverge in higher dimensions).
- There's a way to bring a lot of the above discussions together–and, indeed, syntheses of this sort are the biggest triumphs of 20^{th} century algebraic geometry. If $X \subset \mathbb{P}^2_{\mathbb{Q}}$ is a curve defined over $\mathbb{Q}$, then we can rescale the defining equation of X to have only integer coefficients in a minimal manner (i.e. such that the minimal polynomial is primitive as a trivariate polynomial over $\mathbb{Z}$). It then makes sense to talk about not only $X(\mathbb{Q})$, but in fact $X(k)$ for any field k. Under the additional assumption that $X(\mathbb{C})$ is smooth, it turns out that there is a beautiful relationship between the topology of the curves $X(\mathbb{Q}), X(\mathbb{C})$ and $X(\mathbb{F}_q)$ over different q–this is again brought out in detail by the Weil conjectures, which is a(nother beautiful) story for some other time. Here's a different punchline I want to leave you with. The Mordell-Weil theorem mentioned above asserts that if $X(\mathbb{C})$ has genus 1, then $X(\mathbb{Q})$ is a finitely generated abelian group; this result was shown by Mordell already. Based on this result (and additional considerations), Mordell conjectured in 1922 that if $X(\mathbb{C})$ has genus $g \geq 2$, then, in fact, $X(\mathbb{Q})$ is finite. This fascinating conjecture remained open for a while until it was proven by Faltings in 1983. Isn't this result simply amazing? Somehow, the "rational part" $X(\mathbb{Q})$ of the complex curve $X(\mathbb{C})$ "sees" the topology of the complex curve and decides accordingly whether it wants to be very infinite ($g = 0$), "somewhat infinite" or finitely generated ($g = 1$), or finite ($g \geq 2$).

This is a good ending point for this course. I hope you enjoyed and that it has made you excited to go and learn more algebraic geometry in the future!