An introduction to arithmetic geometry and elliptic curves

Spring 2021

Nate Gillman

Contents

(Lecture 1: January 20, 2021)

1 Overview of algebraic geometry

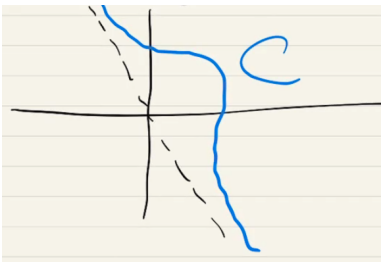
1.1 Introduction

These are my lecture notes for "MATH 2540: Number Theory" taught by Joseph Silverman at Brown University in spring 2021. This was a graduate level topics course which covered elliptic curves, and was conducted entirely via Zoom. For the most part, this course follows his textbook *The Arithmetic of Elliptic Curves,* 2E. At the end of each section in these notes, there is usually a subsection containing the exercises from this text that Joe assigned, as well as my solutions. For most of these problems I collaborated with Veronica Arena, Steven Creech, and Henry Talbott. The purpose of these notes is obviously not to replace Joe's excellent textbook. Rather, I hope they serve as a useful roadmap for someone who wants to chart an expedited and self-contained first pass through the textbook, and that the sections with exercises can serve as a partial solutions manual to aid on this endeavor.

1.2 Diophantine geometry: a whirlwind tour

A *Diophantine equation* is a system of polynomial equations to be solved in $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{F}_q$, $\mathbb{Z}_p$, or some other non-algebraically closed field of number theoretic significance. Examples:

1. $ax + by = c$ is a linear Diophantine equation.
2. Solutions to $x^2 + y^2 = 1$ in $\mathbb{Q}$ characterize Pythagorean triples. The number-theoretic problem of solving this equation over $\mathbb{Q}$ is equivalent to the geometric problem of finding points on the unit circle with coordinates in $\mathbb{Q}$.
3. $x^2 - y^2 = D$, where D is a nonzero integer, for $x, y \in \mathbb{Z}$. As this equation factors as $(x-y)(x+y) = D$, finding integral solutions to this equation is equivalent to factoring D.
4. $x^2 - Ay^2 = 1$, for $A \in \mathbb{Z}_{>0}$ not a square, has infinitely many solutions in $\mathbb{Z}$. In fact, the solutions to this Pell equation form a group. What's actually going on here is that the solutions to the Pell equation are a cyclic subgroup of the group of real points, which follows from the structure of the unit group $\mathbb{Z}[\sqrt{D}]^*$. The moral of this example: we can characterize the number-theoretic solutions by using the group structure on the geometric curve.
5. $x^2 - 2y^3 = D$ looks like this:



This is actually an elliptic curve. On this curve, we consider two different solution sets:

(a) $C_D(\mathbb{Z})$ is finite (this is an ineffective theorem of Thue from 1909; an effective version was proven by Baker in the 1970's, and this won him the Fields medal.)

(b) $C_D(\mathbb{Q})$ is sometimes finite, and sometimes infinite. There is no known algorithm to determine which case a given curve falls in to.

(c) $C_D(\mathbb{R})$ is a group isomorphic to $\mathbb{R}/\mathbb{Z}$, and $C_D(\mathbb{Q})$ is a finitely generated subgroup. In fact, this is true for all elliptic curves.

The idea of *Diophantine geometry:* in order to solve $F_1 = \cdots F_n = 0$ in $\mathbb{Q}$ or $\mathbb{Z}$, first look at the *geometry* (i.e. the solutions in $\mathbb{C}$ or $\mathbb{R}$.) This is algebraic geometry in the classical sense. The goal is to describe the number theory (i.e. the solutions in $\mathbb{Z}$ or $\mathbb{Q}$) in terms of the geometry.

Now we discuss curves. Curves are 1-dimensional objects. *Plane curves* are sets of the form

$$\big\{(x,y) \in \mathbb{A}_k^2 : F(x,y) = 0\big\},$$

where F is some polynomial. If $k = \mathbb{R}$, then this looks like some 1-real-dimensional curve. But we should really be looking at this over $k = \mathbb{C}$. In this case, it's a 1-complex-dimensional curve (which is considered a Riemann surface.) Riemann classified these, in terms of the genus g. Examples:

$F(x,y) = 0$	g	$C(\mathbb{Q})$
$ax + by = c$	0	$\mathbb{Q}$
$x^2 + y^2 = 1$	0	$\mathbb{Q}$
$x^2 - Ay^2 = 1$	0	$\mathbb{Q}$
$x^3 + y^3 = D$	1	finitely generated abelian group
$x^4 + y^4 = D$	3	finite
$y = x^{10}$	0	$\mathbb{Q}$

For example, $x^4 + y^4 = D$ looks like a torus with $g = 3$ holes. The takeaway: given an equation of the form $F(x,y) = 0$, we can associate to the complex solutions (the "geometry") a very coarse geometric invariant: namely, the genus. This is a very coarse measurement of how complicated the geometry is. What do the complex solutions look like? In the genus 0 case, it's basically just $\mathbb{Q}$; in the genus 3 case, there are only finitely many solutions over $\mathbb{Q}$.

But in the $g = 1$ case, the solutions form a finitely generated abelian group (this is known as the Mordell-Weil theorem; we'll prove this later.) The 1-holed tori are known as *elliptic curves.*

Theorem 1.2.1. *Let C be a smooth algebraic curve. Then $C(\mathbb{C})$ is a g-holed torus, and it's fairly easy to compute the genus from the equations defining C.*

Furthermore:

1. *If $g = 0$, then $\mathbb{C}(\mathbb{Q}) \in \{\emptyset, \mathbb{Q}\}$, and we have a good algorithm for determining which it is.*
2. *If $g = 1$ (so C is an elliptic curve) then $C(\mathbb{Q})$ is a finitely generated abelian group, which means $C(\mathbb{Q}) = G_{tors} \times \mathbb{Z}^r$.*
3. *If $g \geq 2$, then $C(\mathbb{Q})$ is finite (Faltings/Vojta)*

Some fun facts about the case $g = 1$: the size of the finite group is ≤ 16 (Mazur, 1970's.) We don't have an algorithm to compute this rank in general, and we don't know if it's bounded or not. Elkies recently found one of rank 28. Proving this rank is bounded is a super important open problem.

The takeaway: the genus of $C(\mathbb{C})$ helps us understand the structure of $C(\mathbb{Q})$. The genus, despite being a very coarse geometric invariant, gives us very interesting qualitative number theoretic descriptions about what the solutions look like. *This is what diophantine geometry is all about.*

(Lecture 2: January 22, 2021)

1.3 Affine varieties

Now we'll present an overview of the algebraic geometry that we'll use. Note: we'll be doing classical algebraic geometry (i.e. we won't be using scheme-theoretic language.) For the rest of the class, we'll assume K is a perfect field, fix an algebraic closure $\overline{K}$ of K, and write $G_K := G_{\overline{K}/K} := \mathrm{Gal}(\overline{K}/K)$ for the absolute Galois group.

Affine space over K is denoted $\mathbb{A}^n := \left\{(x_1, \ldots, x_n) : x_i \in \overline{K}\right\}$, and $\mathbb{A}^n(K) := \{(x_1, \ldots, x_n) : x_i \in K\}$. Galois elements $\sigma \in G_{\overline{K}/K}$ act on $\mathbb{A}^n$ via $P^\sigma = (x_1, \ldots, x_n)^\sigma = (x_1^\sigma, \ldots, x_n^\sigma)$, and $\mathbb{A}^n(K)$ is the space of points fixed under this action. Write $\overline{K}[\mathbf{x}] := \overline{K}[x_1, \ldots, x_n]$. Corresponding to each ideal $I \subseteq \overline{K}[\mathbf{x}]$, we write $V(I) := V_I := \{P \in \mathbb{A}^n : f(P) = 0 (\forall f \in I)\}$ for the algebraic set associated to I. Conversely, given any subset $V \subseteq \mathbb{A}^n$, the associated ideal is $I(V) := I_V := \left\{f \in \overline{K}[\mathbf{x}] : f(P) = 0 (\forall P \in V)\right\}$. Note that the following correspondences hold:

$$V(I(V_0)) = V_0, \qquad I(V(I_0)) = \sqrt{I_0},$$

where $\sqrt{I_0} := \left\{f \in \overline{K}[\mathbf{x}] : f^m \in I_0 \text{ for some } m \geq 1\right\}$. That entire discussion was geometric, as it was over $\overline{K}$. More generally:

Definition 1.3.1. An algebraic set $V \subseteq \mathbb{A}^n$ is *defined over* K if $I(V)$ is generated by polynomials in $K[\mathbf{x}]$.

We write V/K for a variety defined over K. In this case, we have $V(K) = V(\overline{K}) \cap \mathbb{A}^n(K)$. A Galois element $\sigma \in G_K$ acts on $\overline{K}[\mathbf{x}]$ via $f^\sigma = \sum a_{\mathbf{i}}^\sigma \mathbf{x}^{\mathbf{i}}$. This means $f(P)^\sigma = f^\sigma(P^\sigma)$ for $f \in \overline{K}[\mathbf{x}]$. For algebraic sets V defined over K, we have that $V(K) = \left\{P \in V(\overline{K}) : P^\sigma = P (\forall \sigma \in G_K)\right\}$.

Example 1.3.2. Let $V \subseteq \mathbb{A}^2$ be the algebraic set whose ideal is given by $I(V) = (x^d + y^d - 1)$. For even d, $V(\mathbb{R})$ looks like a circle that gets more squashed as d grows. And for $d \geq 3$, Fermat's last theorem says that $V(\mathbb{Q}) = \{(\pm 1, 0), (0, \pm 1)\}$. This example illustrates the interplay of geometry and number theory.

Definition 1.3.3. An algebraic set $V \subseteq \mathbb{A}^n$ is an *affine variety* if $I(V)$ is a prime ideal in $\overline{K}[\mathbf{x}]$.

On the geometry side, this is a desirable property because it's equivalent to V being irreducible. On the algebra side, this is a desirable property because it implies that the coordinate ring is an integral domain. Unpacking this: a polynomial $f \in \overline{K}[\mathbf{x}]$ defines a function $f : V \to \overline{K}$ via evaluation. We have

$$f_1 = f_2 \text{ on } V \iff f_1 - f_2 = 0 \text{ on } V \iff f_1 - f_2 \in I(V) \iff f_1 \equiv f_2 \bmod I(V).$$

Thus, $\overline{K}[\mathbf{x}]/I(V)$ defines functions on V. And if this ring is a nice ring, such as an integral domain, then working with this ring becomes easier.

Definition 1.3.4. If V/K is a variety, then the *affine coordinate ring* of V/K is

$$K[V] := K[\mathbf{x}]/I(V),$$

and the *function field* of V/K is

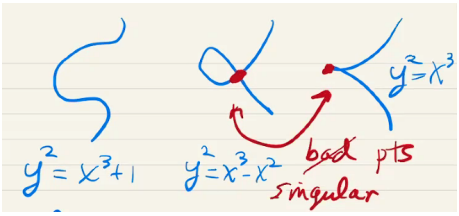
$$K(V) := \mathrm{Frac}(K[V]).$$

Note that G_K acts on $\overline{K}[V]$ and $\overline{K}(V)$, since $f \in I(V/\overline{K})$ implies $f^\sigma \in I(V/\overline{K})$.

Definition 1.3.5. The *dimension* of a variety V is $\dim V := \mathrm{trdeg}\, \overline{K}(V)/\overline{K}$.

This transendence degree is, roughly speaking, the number of independent variables you need to specify the extension $\overline{K}(V)/\overline{K}$. More precisely: there exist $\overline{K}$-algebraically independent quantities $y_1, \ldots, y_d \in \overline{K}(V)$ such that $\overline{K}(V)/\overline{K}(y_1, \ldots, y_d)$ is algebraic. The d here is $\dim V$.

Example 1.3.6. Consider these three curves:



The rightmost curves have points that are "not as nice as the other points on the curve."

How can we algebraically distinguish those points?

Definition 1.3.7. Fix $P \in V \subseteq \mathbb{A}^n$, let $f_1, \ldots, f_m$ generate $I(V)$ in $\overline{K}[\mathbf{x}]$. We say P is *nonsingular* if

$$\operatorname{rank}\left(\frac{\partial f_i}{\partial x_j}(P)\right)_{\substack{1\le i\le m\\ 1\le j\le n}} = n - \dim V.$$

We can reformulate this using algebra. Take $P \in V \subseteq \mathbb{A}^n$. Define

$$M_P := I_{\{P\}}/I_V = \{f \in K[V] : f(P) = 0\}.$$

Then M_P is a maximal ideal, as

$$M_P = \ker(\overline{K}[V] \twoheadrightarrow \overline{K} : f \mapsto f(P)),$$

so $\overline{K}[V]/M_P \cong \overline{K}$.

Proposition 1.3.8. *Let $P \in V \subseteq \mathbb{A}^n$, and $M_P \subseteq \overline{K}[V]$. Then,*

$$\dim_{\overline{K}} M_P/M_P^2 = \dim V \qquad \iff \qquad P \text{ is non-singular.}$$

This algebraic characterization of nonsongularity is easier to check in practice.

Definition 1.3.9. Let $P \in V \subseteq \mathbb{A}^n$. Define $K[V]_P$ to be the localization of $K[V]$ at M_P, so

$$K[V]_P := \left\{\frac{f}{g} \in K(V) : f, g \in K[V], g(P) \neq 0\right\} \subseteq K(V).$$

Algebraically, this is localizing a ring at a maximal ideal. Geometrically, this means we identify functions if they're doing the same thing near to P; more precisely, $\operatorname{Spec}(\overline{K}[V]_P)$ is an affine neighborhood of P in $\operatorname{Spec}(\overline{K}[V]) \cong V$.

(Lecture 3: January 25, 2021)

1.4 Projective varieties

The idea for projective varieties is that affine varieties are "missing points" (e.g: they're not compact, not complete...) A classical definition is $\mathbb{P}^n = \{\text{directions in } \mathbb{A}^{n+1}\}$. A more formal definition:

Definition 1.4.1. *Projective n-space* is $\mathbb{P}^n := (\mathbb{A}^{n+1} \setminus \{0\})/\sim$, where $(a_0, \ldots, a_n) \sim (\lambda a_0, \ldots, \lambda a_n)$ for all $\lambda \in \overline{K}^*$. An equivalence class is written in *homogeneous coordinates* as $[a_0, \ldots, a_n] \in \mathbb{P}^n$.

The K-rational points in projective space are

$$\begin{aligned}\mathbb{P}^n(K) &= \{[a_0,\dots,a_n] : a_i \in K\} \\ &= \left\{[a_0,\dots,a_n] : a_i \in \overline{K},\ \text{some } a_j \neq 0, \text{ every } a_i/a_j \in K\right\} \\ &= \left\{P \in \mathbb{P}^n(\overline{K}) : P^\sigma = P \text{ for all } \sigma \in G_K\right\},\end{aligned}$$

where the last characterization follows from Hilbert's Theorem 90.

Definition 1.4.2. A polynomial $f \in \overline{K}[x_0,\dots,x_n]$ is *homogeneous* of degree d if $f(\lambda x_0,\dots,\lambda x_n) = \lambda^d f(x_0,\dots,x_n)$.

If $P = [a_0,\dots,a_n] \in \mathbb{P}^n$, and f is homogeneous, then $f(P)$ is not well-defined. However, $f(P)$ being zero or nonzero is well-defined.

Definition 1.4.3. A *homogeneous ideal* in $\overline{K}[x_0,\dots,x_n]$ is an ideal I generated by homogeneous polynomials. If I is a homogeneous ideal, then a *projective algebraic set* is

$$V(I) := \left\{P \in \mathbb{P}^n(\overline{K}) : f(P) = 0 \text{ for all homogeneous } f \in I\right\}.$$

If V is a projective algebraic set, then the homogeneous ideal corresponding to V is

$$I(V) := \text{ideal generated by the homogeneous polynomials } f \text{ such that } f(P) = 0 \text{ for all } P \in V.$$

This notation of an algebraic set is well-defined, because homogeneous polynomials generate the ideal, so vanishing at those gives all that is necessary.

Definition 1.4.4. A projective algebraic set V is *defined over* K if $I(V)$ is generated by homogeneous polynomials in $K[x_0,\dots,x_n]$.

Note that the K-points of V are

$$V(K) = V(\overline{K}) \cap \mathbb{P}^n(K) = V(\overline{K})^{G_K},$$

where the latter characterization denotes the fixed points of $V(\overline{K})$ under the action of the absolute Galois group G_K.

Example 1.4.5. Take $K = \mathbb{Q}$. Then points in $\mathbb{P}^n(\mathbb{Q})$ are of the form $[a_0,\dots,a_n]$ with each $a_i \in \mathbb{Q}$.

1. Clearing denominators, we may assume the $a_i \in \mathbb{Z}$.
2. We can then divide by $\gcd(a_0,\dots,a_n)$.

We've shown that every $P \in \mathbb{P}^n(\mathbb{Q})$ has normalized homogeneous coordinates $P = [a_0,\dots,a_n]$, with $a_0,\dots,a_n \in \mathbb{Z}$ and $\gcd(a_0,\dots,a_n) = 1$. These coordinates are unique up to multiplication by ± 1. This will work in general, if we're working in a number field whose ring of integers is a PID; in this case, the normalized homogeneous coordinates are unique up to multiplication by a unit in the ring of integers.

Example 1.4.6. The *Weil height* of $P \in \mathbb{P}^n(\mathbb{Q})$ is $\|P\| := \max\{|a_0|,\dots,|a_n|\}$. One can try to estimate

$$\#\{P \in \mathbb{P}^n(\mathbb{Q}) : \|P\| \leq B\}, \qquad \text{as } B \to \infty.$$

Example 1.4.7. Let $C \subseteq \mathbb{P}^2$ be given by $C : X^2 + Y^2 = 3Z^2$. Then C is defined over $\mathbb{Q}$, as $I(C)$ is generated by the homogeneous polynomial $X^2 + Y^2 - 3Z^2 \in \mathbb{Q}[X,Y,X]$.

We claim that $C(\mathbb{Q}) = \emptyset$.

Proof. Let $[a,b,c] \in C(\mathbb{Q})$, and normalize this point so that $a,b,c \in \mathbb{Z}$ with $\gcd(a,b,c) = 1$. Then $a^2 + b^2 = 3c^2$ implies that $a^2 + b^2 \equiv 0 \pmod 3$, which in turn implies that $a \equiv b \equiv 0 \pmod 3$, since $\left(\frac{-1}{3}\right) = -1$. Then $a^2 + b^2 = 3c^2$ implies $3 \mid c$, which is a contradiction. □

On the other hand, on the curve $C' : X^2 + Y^2 = Z^2$, one can write down an isomorphism $C(\mathbb{Q}) \cong \mathbb{P}^1(\mathbb{Q})$.

Note: reduction mod p is very helpful for showing that a solution doesn't exist.

Definition 1.4.8. A projective algebraic set V is a *projective variety* if its homogeneous ideal $I(V)$ is a prime ideal.

Note that $\mathbb{P}^n$ contains lots of copies of $\mathbb{A}^n$. This is similar to the theory of manifolds, which are built up from many affine-type things. Concretely, for each $0 \leq k \leq n$, we can embed

$$\mathbb{A}^n \hookrightarrow \mathbb{P}^n : (y_1, \ldots, y_n) \mapsto [y_1, \ldots, y_{k-1}, 1, y_k, \ldots, y_n].$$

In the opposite direction, we can map from the compliement of the hyperplane $x_k = 0$:

$$\{P \in \mathbb{P}^n : x_k(P) \neq 0\} \to \mathbb{A}^n : [x_0, \ldots, x_n] \mapsto \left(\frac{x_0}{x_k}, \ldots, \frac{\hat{x_k}}{x_k}, \ldots, \frac{x_n}{x_k}\right).$$

It's clear that projective space is the union of these hyperplanes,

$$\mathbb{P}^n = \bigcup_{k=0}^{n} \{x_k \neq 0\}.$$

Geometrically, this looks like a union of $n+1$ copies of $\mathbb{A}^n$ glued together by transition functions that look like x_i/x_j.

If $V \subseteq \mathbb{P}^n$ is a projective variety, then we can get an affine variety by taking $V^0 = V \cap \mathbb{A}^n$. More precisely, we get $n+1$ affine varieties by intersecting V with the compliment of $n+1$ different hyperplanes, $V^{(i)} = V \cap \{x_i \neq 0\}$. Conversely, given an affine variety $V \subseteq \mathbb{A}^n$, we can take its projective closure $\overline{V} \subseteq \mathbb{P}^n$ as follows. Take coordinates $(y_1, \ldots, y_n)$ for $\mathbb{A}^n$, and embed this in $\mathbb{P}^n$ as $[1, y_1, \ldots, y_n]$. For each polynomial $f \in I(V) \subseteq \overline{K}[y_1, \ldots, y_n]$, let d be the total degree of f (the degree of the monomial of highest degree.) Then we define a *homogenized version* of f as

$$\overline{f}(x_0, \ldots, x_n) := x_0^d f(x_1/x_0, \ldots, x_n/x_0).$$

The projective closure of V is the vanishing locus of all these homogenized polynomials:

Definition 1.4.9. The *projective closure* of an affine algebraic set $V \subseteq \mathbb{A}^n$ is the projective algebraic set

$$\overline{V} := V(\{\overline{f} : f \in I(V)\}) \subseteq \mathbb{P}^n.$$

A super important example:

Example 1.4.10. Consider the curve

$$C : y^2 = x^3 + 2x + 1,$$

and take $f(x, y) = y^2 - x^3 - 2x - 1$, so $\overline{f}(X, Y, Z) = Z^3 f(X/Z, Y/Z) = Y^2 Z - X^3 - 2XZ^2 - Z^3$. Thus the projective closure of C is

$$\overline{C} : Y^2 Z = X^3 + 2XZ^2 + Z^3.$$

The extra points "at infinity" are those with $Z = 0$, which are

$$\overline{C} \setminus C = \{P \in \overline{C} : Z = 0\} = \{[0, Y, 0] \in \mathbb{P}^2\} = \{[0, 1, 0]\}.$$

If $V \subseteq \mathbb{P}^n$ is a projective variety, then choose one $\mathbb{A}^n \subseteq \mathbb{P}^n$ such that $V \cap \mathbb{A}^n \neq \emptyset$. Then $V^0 := V \cap \mathbb{A}^n$ is an affine variety. We define:

1. $\dim V := \dim V^0$ is the *dimension* of V.
2. $\overline{K}(V) = \overline{K}(V^0)$ is the *function field* of V.
3. For $P \in V$, choose i so that $P \in V^{(i)}$. Then $\overline{K}[V]_P := \overline{K}[V^{(i)}]_P$ is the *local ring* of V at P.
4. P is nonsingular on V if P is nonsingular on $V^{(i)}$.

Implicit in this definition: if you take different affine pieces of V_0, then you'll obtain canonically isomorphic function fields, etc. The upshot: in practice, when doing calculations, you can just restrict to the affine that contains the point you're considering.

The function field of $\mathbb{P}^n$ can also be described as

$$\overline{K}(\mathbb{P}^n) = \left\{ \frac{f}{g} : f, g \in K[X_0, \ldots, X_n] \text{ homogeneous of the same degree, } g \neq 0 \right\}.$$

(Lecture 4: January 27, 2021)

1.5 Maps between varieties

Definition 1.5.1. Let $V_1, V_2 \subseteq \mathbb{P}^n$ be projective varieties. A *rational map* $\phi : V_1 \dashrightarrow V_2$ is a map of the form $\phi = [f_0, \ldots, f_m]$, where $f_0, \ldots, f_m \in \overline{K}(V_1)$, not all zero, such that for all $P \in V_1$ for which $f_0, \ldots, f_m$ are defined, we have $[f_0(P), \ldots, f_m(P)] \in V_2$.

In general, a rational function on a variety gives a well-defined function on most of the variety, since there will be just a closed subset of points where it's not well-defined. The only thing stopping a rational function from being defined at a particular point is if all coordinates vanish there. *A crucial point worth highlighting:* the fundamental maps in this setting, rational maps, aren't defined at every point; rather, they're only defined at most points. Note: if $g \in \overline{K}(V_1)^*$, then we can also write $\phi = [gf_0, \ldots, gf_n]$, so there might exist other points on V_1 where ϕ is defined. Thus, it can be tricky to find the maximal subset of V_1 where ϕ is defined.

Definition 1.5.2. A rational map $\phi : V_1 \dashrightarrow V_2$ is *defined at* $P \in V_1$ if:

1. There exists $g \in \overline{K}(V_1)$, generally depending on P, such that $gf_0, \ldots, gf_m \in \overline{K}[V_1]_P$; and
2. For some i, we have $(gf_i)(P) \neq 0$.

Note: the latter condition is equivalent to $gf_i \notin M_P$. In this case, we take $\phi(P) := [(gf_0)(P), \ldots, (gf_m)(P)]$.

Definition 1.5.3. A rational map $\phi : V_1 \dashrightarrow V_2$ is a *morphism* if it is defined at every point of V_1.

Example 1.5.4. Consider the rational map

$$\phi : \mathbb{P}^2 \to \mathbb{P}^2 : [x, y, z] \mapsto [x^2, xy, z^2].$$

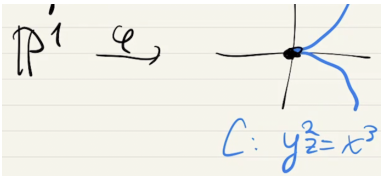
Then ϕ is not a morphism, as it's not defined at $[0, 1, 0]$.

Morphisms behave much more nicely than rational maps, in general.

Definition 1.5.5. A morphism $\phi : V_1 \to V_2$ is an *isomorphism* if it has an inverse $V_2 \to V_1$ that is also a morphism.

There exist bijective morphisms $V_1 \to V_2$ that are not isomorphisms, namely because the inverse map is a rational map which is bijective on all points where it's defined, but isn't defined at all points.

Example 1.5.6. Consider the curve $C : y^2z = x^3$ in $\mathbb{P}^2$. Dehomogenized, this is the cuspidal cubic $y^2 = x^3$:



We define the morphism

$$\psi : \mathbb{P}^1 \to C : [s, t] \mapsto [s^2t, s^3, t^3].$$

Observe that ψ is defined at every point of $\mathbb{P}^1$. However, its inverse is the rational map

$$\psi^{-1} : C \dashrightarrow \mathbb{P}^n : [x, y, z] \mapsto [y, x],$$

which is not defined at the point $[0, 0, 1]$.

Definition 1.5.7. If $\phi : V_1 \to V_2$ is an isomorphism, then V_1 and V_2 are *isomorphic over* K if the morphisms $\phi : V_1 \to V_2$ and $\phi^{-1} : V_2 \to V_1$ are defined over K. In this case, we write $V_1 \cong_K V_2$.

If $\phi : V_1 \xrightarrow{\cong_K} V_2$ is an isomorphism over K, then clearly ϕ induces a bijective map between the K-rational points of V_1 and V_2, so in this case the number theory is basically the same over V_1 and V_2.

Example 1.5.8. Consider the varieties $V_1 : x^2 + y^2 = 1$ and $V_2 : x^2 + y^2 = 3$. Then we have $V_1 \cong_{\mathbb{Q}(\sqrt{3})} V_2$, with isomorphism given by

$$\phi : V_1 \to V_2 : [x, y] \mapsto [\sqrt{3}x, \sqrt{3}y].$$

However, these varieties are not isomorphic over $\mathbb{Q}$, since $V_1(\mathbb{Q}) \cong \mathbb{P}^1(\mathbb{Q})$ (the equation defining $V_1/\mathbb{Q}$ enumerates pythagorean triples) whereas $V_2(\mathbb{Q}) = \emptyset$ (consider the equation modulo 3).

1.6 Exercises

Exercise (Silverman 1.1.a). Let $A, B \in \overline{K}$. Characterize the values of A and B for which the projective variety

$$V_{A,B} : Y^2Z + AXYZ + BYZ^2 = X^3$$

is singular. In particular, as (A, B) ranges over $\mathbb{A}^2$, show that the "singular values" lie on a one-dimensional subset of $\mathbb{A}^2$, so "most" values of (A, B) give a nonsingular variety.

Proof. Define $f(X, Y, Z) = Y^2Z + AXYZ + BYZ^2 - X^3$. We will check in each affine chart whether there are any singular points on $V_{A,B}$.

1. $X \neq 0$: if we normalize by setting $X = 1$, then the Jacobian matrix at $P = (1, Y, Z)$ is

   $$J_P = \left(\tfrac{\partial}{\partial Y} f(1, Y, Z) \quad \tfrac{\partial}{\partial Z} f(1, Y, Z)\right) = \left(2YZ + AZ + BZ^2 \quad Y^2 + AY + 2BYZ\right).$$

   Thus P is a singular point if it satisfies the system

   $$Y^2Z + AYZ + BYZ^2 - 1 = 0, \qquad 2YZ + AZ + BZ^2 = 0, \qquad Y^2 + AY + 2BYZ = 0.$$

   Using Wolfram Alpha (link to computation here) we deduce that $B = A^3/27$, for $A \neq 0$, satisfies the system.

2. $Y \neq 0$: if we normalize by setting $Y = 1$, then the Jacobian matrix at $P = (X, 1, Z)$ is

   $$J_P = \left(\tfrac{\partial}{\partial X} f(X, 1, Z) \quad \tfrac{\partial}{\partial Z} f(X, 1, Z)\right) = \left(AZ - 3X^2 \quad 1 + AX + 2BZ\right).$$

   Thus P is a singular point if it satisfies the system

   $$Z + AXZ + BZ^2 - X^3 = 0, \qquad AZ - 3X^2 = 0, \qquad 1 + AX + 2BZ = 0.$$

   Using Wolfram Alpha (link to computation here) we again obtain the solution $B = A^3/27$ for $A \neq 0$.

3. $Z \neq 0$: if we normalize by setting $Z = 1$, then the Jacobian matrix at $P(X, Y, 1)$ is

   $$J_P = \left(\tfrac{\partial}{\partial X} f(X, Y, 1) \quad \tfrac{\partial}{\partial Y}(X, Y, 1)\right) = \left(AY - 3X^2 \quad 2Y + AX + B\right).$$

   Thus P is a singular point if it satisfies the system

   $$Y^2 + AXY + BY - X^3 = 0, \qquad AY - 3X^2 = 0, \qquad 2Y + AX + B = 0.$$

   Using Wolfram Alpha (link to computation here) we obtain the new solutions $(A, B) = (0, 0)$ and $(A, B) = (\text{nonzero}, 0)$, and we also recover the solution $B = A^3/27$ for $A \neq 0$ from the previous cases.

In summary, the "singular values" (A, B) lie on a union of one-dimensional subspaces of $\mathbb{A}^2$, namely,

$$\text{Sing}(V_{A,B}) = V(A^3 - 27B) \cup V(B),$$

thus generically $V_{A,B}$ is nonsingular. □

Exercise (Silverman 1.2). Find the singular point(s) on each of the following varieties. Sketch $V(\mathbb{R})$.

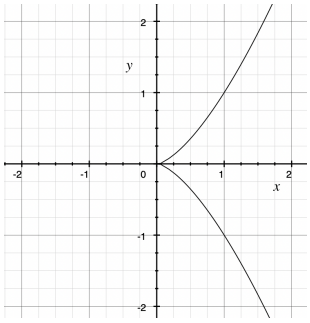
(a) $V : Y^2 = X^3$ in $\mathbb{A}^2$.

(b) $V : 4X^2Y^2 = (X^2 + Y^2)^3$ in $\mathbb{A}^2$.

Solution to (a). If we define $f(X, Y) = Y^2 - X^3$, then the Jacobian matrix at $P = (X, Y)$ is

$$J_P = \left(\tfrac{\partial f}{\partial X}\big|_P \quad \tfrac{\partial f}{\partial Y}\big|_P \right) = \begin{pmatrix} -3X^2 & 2Y \end{pmatrix}.$$

The variety V is singular at P if $\operatorname{rank} J_P = 0$, which happens if and only if $(X, Y) = (0, 0)$. Here is a sketch of $V(\mathbb{R})$:



This confirms that the only singular point is the origin. □

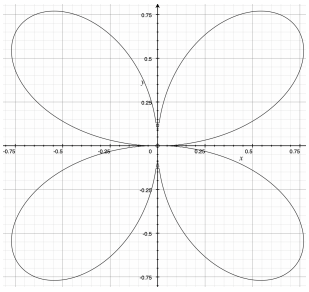
Solution to (b). If we define $f(X, Y) = 4X^2Y^2 - (X^2 + Y^2)^3$, then the Jacobian matrix at $P = (X, Y)$ is

$$J_P = \left(\tfrac{\partial f}{\partial X}\big|_P \quad \tfrac{\partial f}{\partial Y}\big|_P \right) = \begin{pmatrix} 8XY^2 - 6X(X^2 + Y^2)^2 & 8X^2Y - 6Y(X^2 + Y^2)^2 \end{pmatrix}.$$

The variety V is singular at P if $\operatorname{rank} J_P = 0$, which happens if and only if this matrix is identically zero. We restrict our attention to the cases $\operatorname{char} K \notin \{2, 3\}$.

- If $(X, Y) = (0, 0)$, then $J_P = (0\ 0)$, so the origin is a singular point.
- If $X \neq 0$ and $Y = 0$, then
$$\frac{\partial f}{\partial X}\Big|_P = 0 \iff X^2 + Y^2 = 0 \iff X = 0,$$
which is a contradiction, hence P is not a critical point.
- If $X = 0$ and $Y \neq 0$, then by symmetry P is not a critical point.
- If $X, Y \neq 0$, then
$$\frac{\partial f}{\partial X}\Big|_P = 0 \iff \frac{4}{3}Y^2 - (X^2 + Y^2)^2 = 0 \iff \left(\frac{2}{\sqrt{3}}Y - (X^2 + Y^2) \right) \left(\frac{2}{\sqrt{3}}Y + (X^2 + Y^2) \right),$$
and similarly,
$$\frac{\partial f}{\partial Y}\Big|_P = 0 \iff \left(\frac{2}{\sqrt{3}}X - (X^2 + Y^2) \right) \left(\frac{2}{\sqrt{3}}X + (X^2 + Y^2) \right).$$
It follows that the only points of V which kill the Jacobian are those for which $X = Y$ or $X = -Y$. The only points on the variety which satisfy this contraint are given by $4X^4 = (2X^2)^3$, or $0 = x^4(2X^2 - 1)$, which gives the four points
$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \qquad \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right), \qquad \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right), \qquad \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right).$$
But the only points in $\mathbb{A}^2$ which satisfy $X = \pm Y$ that kill the Jacobian have $X = \sqrt{3}$. It follows that there are no singular points with $X, Y \neq 0$.

Here is a sketch of $V(\mathbb{R})$:



This confirms that the only singular point is the origin. □

Exercise (Silverman 1.3). Let $V \subseteq \mathbb{A}^n$ be a variety given by a single equation. Prove that a point $P \in V$ is nonsingular if and only if

$$\dim_{\overline{K}} M_P/M_P^2 = \dim V.$$

Proof. Let $f = 0$ be the equation of V, and define the tangent plane of V at P by

$$T_P := \ker J_P = \left\{ (y_1, \ldots, y_n) \in \mathbb{A}^n : \sum_{i=1}^n \left(\frac{\partial f}{\partial X_i}(P) \right) y_i = 0 \right\}.$$

We will first argue that

$$\psi_P : M_P/M_P^2 \times T_P \to \overline{K} : (g, y) \mapsto \sum_{i=1}^n \left(\frac{\partial g}{\partial X_i}(P) \right) y_i,$$

or equivalently

$$\tilde{\psi}_P : M_P/M_P^2 \to \mathrm{Hom}_{\overline{K}}(T_P, \overline{K}) : g \mapsto \left(y \mapsto \sum_{i=1}^n \left(\frac{\partial g}{\partial X_i}(P) \right) y_i \right), \tag{1.1}$$

is a well-defined perfect pairing of $\overline{K}$-vector spaces. If $g + M_P^2 = g' + M_P^2$, then P is at least a double root of $g - g'$, thus $\frac{\partial (g-g')}{\partial X_i}(P) = 0$. It follows that $(g - g', y) = 0$, so $(g, y) = (g', y)$, which implies that ψ_P is well-defined. If $\tilde{\psi}_P(g)$ is the zero map in $\mathrm{Hom}_{\overline{K}}(T_P, \overline{K})$, then $\sum_{i=1}^n \left(\frac{\partial g}{\partial X_i}(P) \right) y_i = 0$ for every $y \in T_P$, which implies that g has at least a double root at P, so $g = 0$ in M_P/M_P^2. To show that $\tilde{\psi}_P$ is surjective, it suffices to show that for every $y \in T$, there exists some $g \in M_P/M_P^2$ such that $\tilde{\psi}_P(g)(y) = 1$. But if $g \notin M_P^2$, then $\alpha := \tilde{\psi}_P(g)(y) \neq 0$, so $\tilde{\psi}_P(\alpha^{-1} \cdot g)(y) = 1$.

Since (1.1) is a perfect pairing, it follows that $\dim_{\overline{K}} M_P/M_P^2 = \dim_{\overline{K}} T_P = n - \mathrm{rank}\, J_P$. But by definition, P is a nonsingular point of V if and only if $\mathrm{rank}\, J_P = n - \dim V$. Therefore P is nonsingular if and only if $\dim_{\overline{K}} M_P/M_P^2 = \dim V$. □

Exercise (Silverman 1.4). Let $V/\mathbb{Q}$ be the projective variety

$$V : 5X^2 + 6XY + 2Y^2 = 2YZ + Z^2. \tag{1.2}$$

Prove that $V(\mathbb{Q}) = \emptyset$.

Proof. If $(a, b, c) \in V$, then we may assume $a, b, c \in \mathbb{Z}$ and $\gcd(a, b, c) = 1$. We take the equation $5a^2 + 6ab + 2b^2 = 2bc + c^2$ modulo 3, which yields $2a^2 + 2b^2 \equiv 2bc + c^2 \pmod 3$, or equivalently,

$$a^2 + b^2 + c^2 \equiv bc \pmod 3. \tag{1.3}$$

If any two of a, b, c are divisible by 3, then (1.3) implies that the third is as well, which contradicts $\gcd(a, b, c) = 1$. So there are four cases to consider.

- *Case 1:* $a \equiv 0 \pmod 3$. We consider two subcases. If $b \equiv 1 \pmod 3$, then (1.3) implies $c^2 - c + 1 \equiv 0 \pmod 3$, so $c \equiv 2 \pmod 3$. If we write $a = 3a', b = 3b' + 1, c = 3c' + 2$ in (1.2) and reduce modulo 9, then we obtain the contradiction $2 \equiv 1 \pmod 9$. Likewise, if $b \equiv 2 \pmod 3$, then (1.3) implies $c^2 - 2c + 1 \equiv 0 \pmod 3$, so $c \equiv 1 \pmod 3$. If we write $a = 3a', b = 3b' + 2, c = 3c' + 1$ in (1.2) and reduce modulo 9, then we obtain the contradiction $8 \equiv 5 \pmod 9$.
- *Case 2:* $b \equiv 0 \pmod 3$. In this case, (1.3) implies $a^2 + c^2 \equiv 0 \pmod 3$, but because the only quadratic residues modulo 3 are 0 and 1, it must be the case that $a, c \equiv 0 \pmod 3$ as well, which is a contradiction.
- *Case 3:* $c \equiv 0 \pmod 3$. See the previous case.
- *Case 4:* $a, b, c \not\equiv 0 \pmod 3$. In this case, $a^2, b^2, c^2 \equiv 1 \pmod 3$ implies $a^2 + b^2 + c^2 \equiv 0 \pmod 3$, whereas $bc \equiv 1$ or $2 \pmod 3$, which is a contradiction.

It follows that $V(\mathbb{Q}) = \emptyset$. □

Exercise (Silverman 1.6). Let V be the variety

$$V : Y^2Z = X^3 + Z^3.$$

Show that the rational map

$$\phi : V \dashrightarrow \mathbb{P}^2, \qquad \phi = [X^2, XY, Z^2]$$

is a morphism. (Notice that ϕ does not give a morphism $\mathbb{P}^2 \to \mathbb{P}^2$.)

Proof. The rational map ϕ is a priori not defined only at the point $[X, Y, Z] = [0, 1, 0]$. But we can compute that

$$[X^2, XY, Z^2] = [X^4, X^3Y, X^2Z^2] = [XY^2Z - XZ^3, Y^3Z - YZ^3, X^2Z^2] = [XY^2 - XZ^2, Y^3 - YZ^2, X^2Z],$$

which implies that $\phi([0, 1, 0]) = [0, 1, 0]$. As ϕ is defined everywhere, it's a morphism $\phi : V \to \mathbb{P}^2$. □

Exercise (Silverman 1.8). Let $V \subseteq \mathbb{P}^n$ be a variety defined over $\mathbb{F}_q$.

(a) Prove that the q^{th}-power map

$$\phi = [X_0^q, \ldots, X_n^q]$$

is a morphism $\phi : V \to V$. It's called the *Frobenius morphism.*

(b) Prove that ϕ is one-to-one and onto.

(c) If V is irreducible and has positive dimension, prove that ϕ is not an isomorphism.

(d) Prove that $V(\mathbb{F}_q) = \{P \in V : \phi(P) = P\}$.

Proof of (a). As V is defined over $\mathbb{F}_q$, we have that $I(V) = (f_0, \ldots, f_m)$, where the $f_i \in \mathbb{F}_q[X_0, \ldots, X_n]$. Therefore, $[X_0, \ldots, X_n] \in V$ implies $f[X_0, \ldots, X_n] = 0$, thus

$$0 = (f[X_0, \ldots, X_n])^q = f[X_0^q, \ldots, X_n^q].$$

This implies that ϕ is at least a rational map $V \dashrightarrow V$. But in fact, ϕ is only not defined if all of the $X_i^q = 0$ simultaneously, but this only happens if all $X_i = 0$, and this is not a point in projective space so it's not a point on V. □

Proof of (b). If $\phi[X_0, \ldots, X_n] = \phi[Y_0, \ldots, Y_n]$, then $[X_0^q, \ldots, X_n^q] = [Y_0^q, \ldots, Y_n^q]$. Some X_i^q and Y_i^q are nonzero, and we may assume without loss of generality that $X_i^q = Y_i^q = 1$. In the corresponding affine chart $\mathbb{A}^n$, it follows that $X_j^q = Y_j^q \in \overline{\mathbb{F}}_q$ for all $j \neq i$. But because the q'th power map is injective in $\overline{\mathbb{F}}_q$, it follows that $X_i = Y_i$. This implies that ϕ is one-to-one. For surjectivity, given $[Y_0, \ldots, Y_n] \in V$, there exist $X_i \in \overline{\mathbb{F}}_q$ with $X_i^q = Y_i$. And we can compute that $[X_0, \ldots, X_n]$ is indeed a point on V, since

$$(f_i[X_0, \ldots, X_n])^q = f_i[X_0^q, \ldots, X_n^q] = f_i[Y_0, \ldots, Y_n] = 0$$

implies that $f_i[X_0, \ldots, X_n] = 0$. It follows that ϕ is onto. □

Sketch of proof of (c). The morphism $\phi : V \to V$ induces an inclusion

$$\phi^* : \overline{\mathbb{F}}_q(V) \to \overline{\mathbb{F}}_q(V) : f \mapsto f \circ \phi$$

of the function field into itself. It suffices to show that this this inclusion misses some rational function (this is because an isomorphism between varieties induces an isomorphism between their function fields, as the association $\phi \mapsto \phi^*$ is functorial.) But because V is a variety, $V = V(I)$ for some projective prime ideal I. One should use the fact that this ideal is prime in order to demonstrate that there is some rational function which is not the qth power of anything in $\overline{K}(V)$. Namely, we would start out by supposing towards a contradiction that that for every $a/b \in \overline{\mathbb{F}}_q(V)$, there exists $g/h \in \overline{\mathbb{F}}_q(V)$ with

$$a(x_0, \ldots, x_n)h(x_0^q, \ldots, x_n^q) - b(x_0, \ldots, x_n)g(x_0^q, \ldots, x_n^q) \in I(V).$$

□

Proof of (d). By definition, $V(\mathbb{F}_q) = V(\overline{\mathbb{F}}_q) \cap \mathbb{P}^n(\mathbb{F}_q)$. This set is equal to $\{P \in V : \phi(P) = P\}$ because the fixed points of the Frobenius endomorphism $x \mapsto x^q$ on $\overline{\mathbb{F}}_q$ are precisely $\mathbb{F}_q$. □

Exercise (Silverman 1.10). For each prime $p \geq 3$, let $V_p \subseteq \mathbb{P}^2$ be the variety given by the equation

$$V_p : X^2 + Y^2 = pZ^2.$$

(a) Prove that V_p is isomorphic to $\mathbb{P}^1$ over $\mathbb{Q}$ if and only if $p \equiv 1 \pmod 4$.

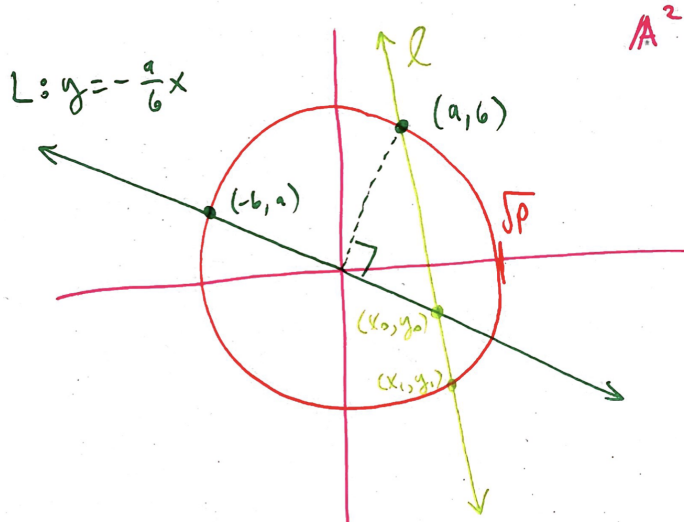
(b) Prove that for $p \equiv 3 \pmod 4$, no two of the V_p's are isomorphic over $\mathbb{Q}$.

Proof of (a). If $p \equiv 3 \pmod 4$, then the Diophantine equation $X^2 + Y^2 = pZ^2$ has no solutions in $\mathbb{Z}$, by the sums of two squares theorem. This implies that $V_p(\mathbb{Q}) = \emptyset$. Next, recall that if $\phi : V \xrightarrow{\cong_K} V'$ is an isomorphism over K, then ϕ induces a bijective map between the K-rational points of V and V'. But since $\mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}$, of course there is no bijective correspondence between $V_p(\mathbb{Q})$ and $\mathbb{P}^1(\mathbb{Q})$, which means V_p and $\mathbb{P}^1$ can't be isomorphic over $\mathbb{Q}$.

Conversely, if $p \equiv 1 \pmod 4$, then by Fermat's theorem on the sum of two squares, we have $p = a^2 + b^2$ for some $a, b \in \mathbb{N}$. Geometrically, this is an integral point on the circle of radius $\sqrt{p}$. Consider the line

$$L : y = -\frac{a}{b}x$$

orthogonal to (a, b). If we take the affine chart $Z = 1$, then it's clear that points on L parameterize those points of $V_p \setminus \{(a, b)\}$, as follows: fix $(x_0, y_0) = (x_0, -ax_0/b) \in L$, draw a new line $\ell(x_0, y_0)$ connecting (a, b) to (x_0, y_0) in $\mathbb{A}^2$, and where it intersects this circle, at (x_1, y_1), is the point that we associate to (x_0, y_0).



If we further associate the point at ∞ to (a,b), then this provides a one-to-one function from V_p to $\mathbb{P}^1$. In order to show that this correspondence is a morphism, we will explicitly provide the rational map.

The equation for $\ell(x_0,y_0)$ is

$$\ell(x_0,x_1): y-y_0=\frac{b-y_0}{a-x_0}(x-x_0).$$

A tedious calculation (link here for computation of x_1) reveals that this line intersects the circle $x^2+y^2=p$ at the point

$$(x_1,y_1)=\left(\frac{a(x_0^2-b^2)+2b^2x_0}{b^2+x_0^2},\frac{b^2+ax_0}{ab-bx_0}\left[\frac{a(x_0^2-b^2)+2b^2x_0-x_0(b^2+x_0^2)}{b^2+x_0^2}\right]-\frac{a}{b}x_0.\right) \tag{1.4}$$

This function $\mathbb{A}^1\to\{[X,Y,Z]\in V_p: Z=1\}: x_0\mapsto(x_1,y_1)$ extends to a rational map $\mathbb{P}^1\dashrightarrow V_p$ as

$$\begin{aligned}\psi:&\mathbb{P}^1\dashrightarrow V_p:[P,Q]\mapsto\\&\left[\frac{a(P^2-b^2Q^2)+2b^2PQ}{b^2Q^2+P^2},\frac{b^2Q+aP}{abP-bQ}\left[\frac{a(P^2Q-b^2Q^3)+2b^2PQ^2-P(b^2Q^2+P^2)}{b^2Q^3+P^2Q}\right]-\frac{aP}{bQ},Q^2\right],\end{aligned}$$

which is obtained from (1.4) by homogenizing via the variable transformation $x_0=P/Q$. If we wanted, we could clear denominators and get a rational map of degree 8.

The inverse map is constructed in a similar fashion: given a point (x_1,y_1) on the circle $x^2+y^2=p$, we connect (a,b) to (x_1,y_1) via a line $\ell'(x_1,y_1)$, and where this line intersects $L: y=-ax/b$, at (x_0,y_0), is the point that we associate to (x_1,y_1). The equation for $\ell'(x_1,y_1)$ is

$$\ell'(x_1,y_1): y-b=\frac{b-y_1}{a-x_1}(x-a),$$

and a straightforward calculation reveals that $\ell'(x_1,y_1)$ intersects L at the point

$$(x_0,y_0)=\left(\frac{b^2c-abd}{p-bd-ac},\frac{b-y_1}{a-x_1}\left(\frac{b^2c-abd}{p-bd-ac}-a\right)+b\right).$$

This function $\{[X,Y,Z]\in V_p: Z=1\}\to\mathbb{A}^1:(x_1,y_1)\mapsto(x_0,y_0)$ extends to a rational map

$$\phi: V_p\dashrightarrow\mathbb{P}^1$$

in the obvious way. By construction the rational functions ψ and ϕ are inverses on their respective domains, so it remains to argue that these functions are defined everywhere. This appears to be a very messy computation, so we'll leave that part unfinished. □

Proof of (b). If V and V' are isomorphic over $\mathbb{Q}$, then they're isomorphic over $\mathbb{Z}$, so they're isomorphic over $\mathbb{F}_p$, which implies that $\#V(\mathbb{F}_p)=\#V'(\mathbb{F}_p)$. Therefore, in order to show that V_p and V_q are not isomorphic over $\mathbb{Q}$, it suffices to show that $\#V_p(\mathbb{F}_p)\neq\#V_q(\mathbb{F}_p)$.

1. We have that $V_p(\mathbb{F}_p)=\{[0,0,1]\}$ because $p\equiv 3\pmod 4$ implies that there are no nontrivial solutions to $x^2+y^2\equiv 0\pmod p$.
2. We have that $\#V_q(\mathbb{F}_p)=p+1$ because $p,q\equiv 3\pmod 4$ implies that there exist $p+1$ solutions in $\mathbb{P}^2_{\mathbb{F}_p}$ to $x^2+y^2\equiv qz^2\pmod p$ (in the finite field $\mathbb{F}_p$, we can solve $x^2+y^2\equiv a\pmod p$ for any $a\in\mathbb{F}_p$.)

This completes the proof. □

2 Algebraic curves

In this class, by *curves* we mean 1-dimensional projective varieties. We'll spend the semester studying one very special kind of curve... but before then, we'll look at the theory of curves in general.

2.1 Notation

- C/K will be a nonsingular projective curve defined over K.
- We'll write $C = C(\overline{K})$.
- $\overline{K}(C)$ (resp. $K(C)$) is the field of rational functions on C with coefficients in $\overline{K}$ (resp. K).
- $K[C]_P$ is the local ring of rational function defined at P.
- $M_P := \{f/g \in K[C]_P : f(P) = 0, g(P) \neq 0\}$ is the maximal ideal in $K[C]_P$.

2.2 Basic notions

Proposition 2.2.1. *Let P be a smooth point of C. Then the local ring $(\overline{K}[C]_P, M_P)$ is a discrete valuation ring.*

Proof. P is nonsingular means $\dim_{\overline{K}} M_P/M_P^2 = 1$, but really we should think of $\overline{K}$ as $\overline{K}[C]_P/M_P$. In other words, $(R, M) = (\overline{K}[C]_P, M_P)$ is a local ring with the property that $\dim_{R/M}(M/M^2) = 1$. This condition on a local ring is equivalent to R being a discrete valuation ring (so long as R is Noetherian, which is true, as all rings we consider are quotients of polynomial rings over fields). □

Notation 2.2.2. The normalized valuation is denoted $\operatorname{ord}_P : \overline{K}[C]_P \to \mathbb{N}$.

The valuation ord_P is "normalized" in the sense that it surjects onto $\mathbb{N}$. We should think of $\operatorname{ord}_P(f)$ as the order of vanishing of f at P. Importantly, we can extend the normalized valuation as follows:

1. The normalized valuation on $\overline{K}[C]_P$ is $\operatorname{ord}_P : \overline{K}[C]_P \to \mathbb{N}$.
2. The valuation on $\overline{K}[C]_P$ extends to a homomorphism $\operatorname{ord}_P : \overline{K}(C)^* \to \mathbb{Z}$.
3. We extend this latter extension to $\operatorname{ord}_P : \overline{K}(C) \to \mathbb{Z} \cup \{\infty\}$ via $\operatorname{ord}_P(0) = \infty$

Definition 2.2.3. A *uniformizer* for C at P is a function $t \in \overline{K}(C)$ with $\operatorname{ord}_P(t) = 1$.

If $\operatorname{ord}_P(f) > 0$, we say f *vanishes at* P, and if $\operatorname{ord}_P(f) < 0$, we say f has a *pole at* P. In this terminology, we think of a uniformizer as a function which vanishes at P to order 1. There will be lots of them, of course.

Example 2.2.4. Take $C = \mathbb{P}^1$. The rational functions on $\mathbb{P}^n$ are of the form $f = h/g$, where h and g are homogeneous polynomials of the same degree. In this case, we can factor

$$f = \frac{g}{h} = \frac{\prod(\alpha_i X - \beta_i Y)^{m_i}}{\prod(\gamma_i X - \delta_i Y)^{n_i}},$$

where $\sum m_i = \sum n_i$. In other words, the total number of zeros equals the total number of poles. Further, the normalized valuation for f in this case is

$$\operatorname{ord}_P f = \begin{cases} m_i & P = [\beta_i, \alpha_i] \\ -n_i & P = [\delta_i, \gamma_i] \\ 0 & \text{else.} \end{cases}$$

Proposition 2.2.5. *For a nonzero rational function $f \in \overline{K}(C)^*$, the following are true:*

(a) f has finitely many zeros and poles.

(b) The total number of zeros equals the total number of poles. More precisely,

$$\sum_{\substack{P \in C \\ \operatorname{ord}_P(f) > 0}} \operatorname{ord}_P(f) = \sum_{\substack{P \in C \\ \operatorname{ord}_P(f) < 0}} -\operatorname{ord}_P(f).$$

Idea of proof: reduce to the case of $\mathbb{P}^1$ by mapping C onto $\mathbb{P}^1$.

Proposition 2.2.6. *Let C be a smooth curve, $f \in \overline{K}(C)$, and suppose f has no zeros or poles. Then in fact $f \in \overline{K}^*$, i.e., f is constant.*

This is a version of Liouville's theorem from complex analysis. The point here is that the projective curve $C(\mathbb{C})$ is compact (as it's a g-holed torus) so a holomorphic function on this compact Riemann surface with no poles is necessarily constant.

2.3 Maps between curves

Proposition 2.3.1. *Let C be a smooth projective curve, $V \subseteq \mathbb{P}^N$ a projective variety, and $\phi : C \dashrightarrow V$ a rational map. Then ϕ is a morphism.*

Proof. Write $\phi = [f_0, \dots, f_N]$ with each $f_i \in \overline{K}(C)$. Take $P \in C$, and let t be a uniformizer at P, so by definition $t \in \overline{K}[C]_P$ and $\operatorname{ord}_P(t) = 1$. (Here we're using the hypothesis that C is smooth at P, as this implies that $\overline{K}[C]_P$ is a discrete valuation ring, which is how we get the uniformizer.) Let $m_i = \operatorname{ord}_P(f_i)$. Because $\overline{K}[C]_P$ is a discrete valuation ring, we can multiply ϕ by an appropriate power of t to kill all the poles, but not so much. That is, we'll consider $[t^m f_0, \dots, t^m f_N]$.

What value should we choose for m? We want all of the $t^m f_i$ to be in $\overline{K}[C]_P$, and we want at least one to not be in M_P. Equivalently, we want all them to have $\operatorname{ord}_P(t^m f_i) \geq 0$, and we want some $\operatorname{ord}_P(t^m f_i) = 0$. We know $\operatorname{ord}_P(t^m f) = m + \operatorname{ord}_P(f)$, so we want $m + m_i \geq 0$ for all i, and $m + m_i = 0$ for some i. So we should take $m = -\min\{m_i\} = \max\{-m_i\}$. □

In particular, if C is a smooth curve and $f \in \overline{K}(C)$, then a priori f is just a rational map, so at some points of C, f might not take a finite complex value; however, this proposition says that f is actually a morphism. Naturally, it's the morphism $[f, 1]$. If we think of f as this map between curves, then (by an abuse of notation) we write

$$f : C \to \mathbb{P}^1 : f(P) = \begin{cases} [f(P), 1] & f \in \overline{K}[C]_P \\ [1, 0] & f \notin \overline{K}[C]_P. \end{cases}$$

In other words, *we've extended the rational map $f : C \dashrightarrow \mathbb{C}$ by mapping those points which take value ∞ to that value $\infty = [1, 0] \in \mathbb{P}^1$.* Conversely, if $\phi : C \to \mathbb{P}^1$ is given by $\phi = [f_0, f_1]$, then we identify ϕ with $f_0/f_1 \in \overline{K}(C)$. In order words, *rational functions and morphisms to $\mathbb{P}^1$ are basically the same for smooth curves.*

(Lecture 5: January 29, 2021)

Today we'll provide an overview of the theory of maps between curves, mostly without proof.

Theorem 2.3.2. *Let C_1, C_2 be smooth curves over K, and $\phi : C_1 \to C_2$ a morphism over K. Then ϕ is either constant or surjective.*

In this context, "surjectivity" means surjectivity from the K-points $C_1(K)$ onto the K-points $C_2(K)$. This theorem essentially comes from compactness. The complex analysis version of this result: these curves are Riemann surfaces, and ϕ is an open map between them, so it must be surjective.

Theorem 2.3.3. *The morphism $\phi : C_1 \to C_2$ induces a map on the function fields*

$$\phi^* : \overline{K}(C_2) \to \overline{K}(C_1) : \phi^* f = f \circ \phi.$$

As field homomorphisms are embeddings, the previous result says that $\overline{K}(C_2)$ is actually a subfield of $K(C_1)$. Correspondingly:

Theorem 2.3.4. *If the morphism $\phi : C_1 \to C_2$ is nonconstant, then:*

(a) The index $[\overline{K}(C_1) : \phi^* \overline{K}(C_2)]$ is finite, and we define $\deg \phi := [\overline{K}(C_1) : \phi^* \overline{K}(C_2)]$ to be the degree of ϕ.

(b) Given an inclusion $i : K(C_2) \hookrightarrow K(C_1)$, there exists a unique nonconstant morphism $\phi : C_1 \to C_2$ such that $i = \phi^*$.

(c) Given C_1/K and a subfield $L \subseteq K(C_1)$ of finite index, then there exists a unique (up to isomorphism) smooth projective curve C_2 and $\phi : C_1 \to C_2$ such that $L = \phi^*\overline{K}(C_2)$.

Example 2.3.5. A map $\phi : \mathbb{P}^1 \to \mathbb{P}^1$ is of the form $\phi = [F, G]$, where F and G are homogeneous polynomials with no common roots. In this case, $\deg \phi = \deg F = \deg G$ according to the definition in (a), as

$$\deg \phi = [\overline{K}(\mathbb{P}^1), \phi^*\overline{K}(\mathbb{P}^1)] = [\overline{K}(t) : \phi^*\overline{K}(t)] = \left[\, \overline{K}(t) : \left\{g(\phi)/h(\phi) : g, h \in \overline{K}[t]\right\}\right].$$

2.4 Ramification

Let $\phi : C_1 \to C_2$ be a morphism between smooth projective curves. For almost all $Q \in C_2$, we basically have $\deg \phi = \#\phi^{-1}(Q)$. To precisely state this result, we need the following definition.

Definition 2.4.1. The morphism $\phi : C^1 \to C^2$ is *separable*, or *inseparable*, or *totally inseparable*, if its corresponding field extension $\overline{K}(C_1)/\phi^*\overline{K}(C_2)$ has that property.

1. We write $\deg_s(\phi)$ for the *separable degree*.
2. We write $\deg_i(\phi)$ for the *inseparable degree*.

Recall that an extension of fields is *separable* if every irreducible polynomial, when factored completely, has distinct roots. An extension is *inseparable* if it is not separable, and an extension is *totally inseparable* if for every element, its irreducible polynomial has only one root. (Recall that if $\operatorname{char} K = 0$ every field is separable.) To complete the discussion above: one can show that for all but finitely many $Q \in C_2$, we have $\#\phi^{-1}(Q) = \deg_s \phi$.

Theorem 2.4.2. *A morphism $\phi : C_1 \to C_2$ is an isomorphism if and only if $\deg \phi = 1$.*

Informally speaking, a *ramification point* is where you get fewer points in the preimage than you should because of things like tangency. Formal definitions use algebra:

Definition 2.4.3. Let $\phi : C_1 \to C_2$ be a morphism of smooth projective curves. Fix $P \in C_1$ and choose a uniformizer $t_P \in K(C_1)$ at P, as well as a uniformizer $t_{\phi(P)} \in K(C_2)$ at $\phi(P)$. Its pullback $\phi^* t_{\phi(P)}$ vanishes at P, since by definition $(\phi^* t_{\phi(P)})(P) = t_{\phi(P)} \circ \phi(P) = 0$. The order of vanishing of this pullback is the *ramification index* of ϕ at P, and it is denoted

$$e_\phi(P) := \operatorname{ord}_P(\phi^* t_{\phi(P)}).$$

We say P is *ramified* (resp. *unramified*) at P if it $e_\phi(P) > 1$ (resp. $e_\phi(P) = 1$).

It's clear that this definition is well-defined regardless of the choices of uniformizer. Roughly speaking, the ramification index $e_\phi(P)$ tells you how many times you should count points in the preimage of $\phi(P)$. If the index is greater than one, then P occurs in the preimage of $\phi(P)$ with multiplicity greater than 1.

Theorem 2.4.4. *Let $\phi : C_1 \twoheadrightarrow C_2$ be a surjective morphism.*

(a) If $Q \in C_2$, then

$$\sum_{P \in \phi^{-1}(Q)} e_\phi(P) = \deg \phi.$$

(b) The set $\left\{Q \in C_2 : \#\phi^{-1}(Q) \neq \deg \phi\right\}$ is finite. Equivalently, the set $\{P \in C_1 : e_\phi(P) \geq 2\}$ is finite.

(c) Given maps $C_1 \xrightarrow{\phi} C_2 \xrightarrow{\psi} C_3$, we have $e_{\psi\phi}(P) = e_\phi(P) e_\psi(\phi P)$ for every $P \in C_1$.

Part (a) looks similar to the result that $\sum e(\mathfrak{p}_i) f(\mathfrak{p}_i) = [L : K]$. Both of these results are special cases of the same theorem about how prime ideals split in Dedekind domains. Part (b) is the analog of the fact that in a number field, only finitely many primes ramify. And part (c) is the analog of the fact that ramification indices multiply in towers of fields, but translated into this geometric language.

2.5 The Frobenius map

Suppose K is a perfect field (meaning it doesn't have any inseparable extensions) with $\operatorname{char} K = p$ (for example, K a finite field). Let $q = p^r$.

Definition 2.5.1. If we write $f(\mathbf{x}) \in K[\mathbf{x}]$ as $f(\mathbf{x}) := \sum_i \mathbf{a}_i \mathbf{x}^{\mathbf{i}}$, then we let[1]

$$f^{(q)}(\mathbf{x}) := \sum_i \mathbf{a}_i^q \mathbf{x}^{\mathbf{i}}.$$

This is simply a polynomial where we've raised each coefficient to the q'th power.

Definition 2.5.2. If C/K is a curve, then we define

$$C^{(q)} := V(I(C)^{(q)}).$$

Definition 2.5.3. The *q-power Frobenius map* is

$$F_q : C \to C^{(q)} : [x_0, \dots, x_n] \mapsto [x_0^q, \dots, x_n^q].$$

Example 2.5.4. Fix $A, B \in K$, and suppose

$$C : Y^2Z = X^3 + AXZ^2 + BZ^3$$

is a curve in $\mathbb{P}^2$. Then

$$C^{(q)} : Y^2Z = X^3 + A^qXZ^2 + B^qZ^3.$$

Furthermore, we have $F_q([X, Y, Z]) = [X^q, Y^q, Z^q]$ is on $C^{(q)}$ because

$$Y^{2q}Z^q - X^{3q} - A^qX^qZ^{2q} - B^qZ^{3q} = (Y^2Z - X^3 - AXZ^2 - BZ^3)^q,$$

due to the fact that $\operatorname{char} K = p$.

If we're working over an algebraically closed field, then $F_q : C \to C^{(q)}$ is surjective, but it will be ramified everywhere. In fact, F_q is totally inseparable. Furthermore, $\deg F_q = \deg_i F_q$, whereas $\deg_s F_q = 1$. The example to keep in mind:

Example 2.5.5. Consider $F_p : \mathbb{P}^1 \to \mathbb{P}^1$ over $\overline{\mathbb{F}}_p$. Then the pullback is

$$F_p^* \overline{\mathbb{F}}_p(T) = \overline{\mathbb{F}}_p(T) \circ F_p = \overline{\mathbb{F}}_p(T^p),$$

and notice that the extension of fields has degree

$$[\mathbb{F}_p(T) : \mathbb{F}_p(T^p)] = p$$

and is totally inseparable, since the minimal polynomial for this extension is $X^p - T^p$, which is irreducible in $\mathbb{F}_p(T^p)[x]$, and over the splitting field it only has one root.

Recall the theorem from algebra:

Theorem 2.5.6. *If L/K is a field extension, then there exists an intermediate field E such that L/E is totally inseparable and E/K is separable.*

The consequence of this is that maps $C_1 \to C_2$ always factors as

$$C_1 \xrightarrow{F_q} C_1^{(q)} \xrightarrow{separable} C_2.$$

This basically means that even in characteristic p, we can study problems by studying separable maps (which are easier to work with) and Frobenius maps (which we have a simple description for).

[1] Here we're using multiexponents, i.e., $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{a}_i^q \mathbf{x}^{\mathbf{i}} = a_{i_1 \cdots, i_n}^q x_1^{i_1} \cdots x_n^{i_n}$.

2.6 Divisors on curves

Given a function $f \in K(C)$, one can easily list its zeros and poles. We want to go the opposite direction: namely, we'd like to specify zeros/poles, and then construct a function that has those zeros/poles. To do this, we'll abstract away from actual functions, and instead discuss "points with multiplicities." This is the motivation for defining divisors.

Definition 2.6.1. Let C/K be a smooth curve. A *divisor* on C is a formal sum

$$\sum_{P \in C} n_P(P),$$

where the $n_P \in \mathbb{Z}$, and only finitely many n_P are nonzero. The *divisor group* of C is

$$\mathrm{Div}(C) := \left\{ \text{divisors } D = \sum_{P \in C} n_P(P) \right\},$$

which is simply the free abelian group generated by the points of C. The *degree* of a divisor is

$$\deg \sum_{P \in C} n_P(P) := \sum_{P \in C} n_P.$$

The *divisors of degree zero* are

$$\mathrm{Div}^0(C) := \{D \in \mathrm{Div}(C) : \deg D = 0\}.$$

(Lecture 6: February 1, 2021)

Definition 2.6.2. The *divisors on C defined over K* are

$$\mathrm{Div}_K(C) = \{D \in \mathrm{Div}(C) : D^\sigma = D (\forall \sigma \in G_K)\},$$

where the Galois action is $(\sum n_P(P))^\sigma := \sum n_p(P^\sigma)$.

Note that $(\mathrm{Div}(f))^\sigma = \mathrm{Div}(f^\sigma)$, since σ acting on f will permute roots and poles.

Definition 2.6.3. The *divisor* of $f \in \overline{K}(C)$ is

$$(f) := \mathrm{Div}(f) := \sum_{P \in C} \mathrm{ord}_P(f) P.$$

Interestingly, if we think of taking divisors as a function

$$\mathrm{Div} : \overline{K}(C)^* \to \mathrm{Div}(C),$$

then this is actually a homomorphism, which follows from the fact that ord_P is a valuation. In the number field setting, this corresponds to mapping an element of the field to its corresponding fractional ideal. This justifies the following terminology:

Definition 2.6.4. A divisor $D \in \mathrm{Div}(C)$ is a *principal divisor* if $D = (f)$ for some $f \in \overline{K}(C)^*$.

1. We say D_1 is *linearly equivalent* to D_2, and write $D_1 \sim D_2$, if $D_1 - D_2$ is principal.
2. The *Picard group* is $\mathrm{Pic}(C) := \mathrm{Div}(C)/\sim$.
3. We let $\mathrm{Pic}_K(C) = \{[D] : [D^\sigma] = [D] \forall \sigma \in G_K\}$, where $[D]$ denotes a divisor class.

This last condition means that for all σ there exists $f_\sigma \in \overline{K}(C)^*$ such that $D^\sigma = D + (f_\sigma)$.

Proposition 2.6.5. *Let $f \in \overline{K}(C)^*$.*

(a) $\mathrm{Div}(f) = 0$ *if and only if* $f \in \overline{K}^*$ *(i.e. every nonconstant function has some zeros or poles)*

(b) $\deg(\mathrm{Div}(f)) = 0$ *(i.e. the number of zeroes equals the number of poles)*

In particular, there is a short exact sequence

$$1 \to \overline{K}^* \to \overline{K}(C)^* \xrightarrow{\mathrm{Div}} \mathrm{Div}^0(C) \to \mathrm{Pic}^0(C) \to 0.$$

This corresponds to a short exact sequence from algebraic number theory,

$$1 \to \text{units in } K \to K^* \to \text{fractional ideals of } K \to \text{ideal class group of } K \to 1.$$

In fact, the fields $\overline{K}^*$ and $\overline{K}(C)^*$ are the fraction fields of Dedekind domains; this explains the analogy.

Example 2.6.6. The map deg is a surjective homomorphism $\mathrm{Div}(\mathbb{P}^1) \to \mathbb{Z}$, with kernel given by $\mathrm{Div}^0(C)$. In fact, we have that $\mathrm{Div}^0(C)$ is exactly the space of principal divisors on $\mathbb{P}^1$ (see example 3.2 on Silverman pp. 28) so in this case, deg induces an isomorphism $\mathrm{Pic}(\mathbb{P}^1) \xrightarrow{\cong} \mathbb{Z}$.

A map $\phi : C_1 \to C_2$ induces a map on divisor groups (in the opposite direction) which respects linear equivalence. This map is defined to be

$$\phi^* : \mathrm{Div}(C_2) \to \mathrm{Div}(C_1) : \sum_{P \in C_2} n_P(P) \mapsto \sum_{P \in C_2} n_P \sum_{Q \in \phi^{-1}(P)} e_\phi(Q)(Q).$$

In words, we we just take the preimages, but we take them with the approproate multiplicity (which is necessary if there is ramification.)

Example 2.6.7. Recall that a nonzero rational function $f \in \overline{K}(C)^*$ gives a map

$$f : C \to \mathbb{P}^1 : P \mapsto \begin{cases} [f(P), 1] & f \text{ doesn't have a pole at } P \\ [1, 0] & f \text{ has a pole at } P. \end{cases}$$

Then $\mathrm{div}(f)$ is simply the zeros minus the poles, so it's the pullback

$$\mathrm{div}(f) = f^*((0) - (\infty)) = \sum_{Q \in f^{-1}(0)} e_f(Q) \cdot Q - \sum_{Q \in f^{-1}(\infty)} e_f(Q) \cdot Q.$$

We can also give a map in the other direction,

$$\phi_* : \mathrm{Div}(C_1) \to \mathrm{Div}(C_2) : \sum_{P \in C_1} n_P(P) \mapsto \sum_{P \in C_1} n_P(\phi(P)).$$

Proposition 2.6.8. *Given a map* $\phi : C_1 \to C_2$ *between smooth curves, we have*

$$\phi_* \phi^*(D) = (\deg \phi) D$$

for every divisor D*.*

Proof. We compute

$$\begin{aligned} \phi_* \phi^* \left(\sum_P n_P(P) \right) &= \phi_* \left(\sum_P n_P \sum_{Q \in \phi^{-1}(P)} e_\phi(Q)(Q) \right) \\ &= \sum_P n_P \sum_{Q \in \phi^{-1}(P)} e_\phi(Q)(\phi(Q)) \\ &= \sum_P n_P \left(\sum_{Q \in \phi^{-1}(P)} e_\phi(Q) \right) (P) \\ &= (\deg \phi) \sum_P n_P(P). \end{aligned}$$

□

Divisors will be very important because they're used to define functions.

Question 2.6.9. A function f gives a divisor $\mathrm{div}(f)$. When does a divisor D give a function f with $\mathrm{div}(f) = D$?

The answer is provided by the Riemann Roch theorem, but to state this very important result we'll need differentials.

2.7 Differentials

Definition 2.7.1. The space of *(meromorphic) differential* 1*-forms* is denoted Ω_C, and it's the $\overline{K}(C)$-vector space spanned by the symbols $\{dx : x \in \overline{K}(C)\}$ subject to the the relations

1. $d(x+y) = dx + dy$
2. $d(xy) = xdy + ydx$
3. $dc = 0$ for all $c \in \overline{K}$

We'll see later that the space of holomorphic differentials will have a similar definition, but they will be a $\overline{K}$-vector space. Differentials are supposed to "linearize" maps in some sense. A nonconstant map $\phi : C_1 \to C_2$ between smooth projective curves gives rise to an associated pullback map on the function field

$$\phi^* : \overline{K}(C_2) \to \overline{K}(C_1) : f \mapsto f \circ \phi,$$

which in turn induces the map

$$\phi^* : \Omega_{C_2} \to \Omega_{C_1} : \sum_{i=1}^{n} f_i dx_i \mapsto \sum_{i=1}^{n} (\phi^* f_i) d(\phi^* x_i), \qquad \text{where } f_i, x_i \in \overline{K}(C)^*.$$

Proposition 2.7.2. *(a) Ω_C is a 1-dimensional $\overline{K}(C)$-vector space.*

(b) $\phi : C_1 \to C_2$ is separable if and only if $\phi^* : \Omega_{C_2} \to \Omega_{C_1}$ is non-zero.

Proof. For (a), given $x, y \in \overline{K}(C) \setminus \overline{K}$, we claim that dx and dy are $\overline{K}(C)$-linearly dependent. This will imply that $\dim_{\overline{K}(C)} \Omega_C \leq 1$. Consider the tower

$$\overline{K}(C) \supseteq \overline{K}(x) \supseteq \overline{K}.$$

The right tower has transcendence degree at least 1, and the whole tower has transcendence degree 1, thus the left tower is an algebraic extension. Thus, as $y \in K(C)$, there exists a polynomial $F(T) \in \overline{K}(x)[T]$ such that $F(y) = 0$. If we write the coefficients as rational functions in x and clear the denominators, this means there exists a polynomial $G(S,T) \in \overline{K}[S,T]$ with $G(x,y) = 0$. This implies that $G_S(x,y)dx + G_T(x,y)dy = 0$ (this follows formally from the definition of differentials; to convince oneself of this, first consider the case $G = S^n T^m$) and this in turn implies that dx and dy are linearly dependent over the function field $\overline{K}(C)$. □

Let C be a curve, $P \in C$, and $t \in \overline{K}(C)$ a uniformizer ar P. For any differential form $\omega \in \Omega_C$, one can show that there exists a unique $g \in \overline{K}(C)$ such that $\omega = gdt$, and we'll write $g = \omega/dt$. In particular, for any $f \in \overline{K}(C)$, then we can look at the differential df, which is some multiple of dt, so $df/dt \in \overline{K}(C)$ is well-defined.

Claim 2.7.3. If f is regular at P (meaning it has no zeros or poles) then df/dt is also regular at P.

Definition 2.7.4. Let $\omega \in \Omega_C$ be a nonzero meromorphic 1-form, and t a uniformizer at P. Then $\omega/dt \in \overline{K}(C)^*$, so we define the *order* of the differential ω at the point P to be

$$\mathrm{ord}_P(\omega) := \mathrm{ord}_P(\omega/dt).$$

A meromorphic 1-form has only finitely many zeros and poles, under this definition. Namely:

Claim 2.7.5. For all but finitely many $P \in C$, we have $\operatorname{ord}_P(\Omega) = 0$.

Definition 2.7.6. The *divisor* of a differential form ω on C is

$$\operatorname{div}(\omega) := \sum_{P \in C} \operatorname{ord}_P(\omega)(P).$$

It is definitely not the case that differentials have the same number of poles and zeros, as this next example demonstrates.

Example 2.7.7. Let $C = \mathbb{P}^1$ with coordinates $[x, y]$. Consider the rational function $t = x/y \in \overline{K}(C)$ and the differential form $\omega = dt$. Our goal is to compute $\operatorname{ord}_P(\omega)$. There are two cases to consider.

1. If $P = [\alpha, 1]$ with $\alpha \neq 0$, then a uniformizer at P is given by $t_P = t - \alpha$. Note that $t_P([\alpha, 1]) = (\alpha/1) - \alpha = 0$, so t_P vanishes at P to order 1. But $dt_P = d(t - \alpha) = dt$, so we can compute
$$\operatorname{ord}_{[\alpha,1]}(dt) := \operatorname{ord}_{[\alpha,1]}(dt/dt_P) = \operatorname{ord}_{[\alpha,1]}(dt/dt) = \operatorname{ord}_{[\alpha,1]}(1) = 0.$$
2. If $P = [1, 0]$, i.e. the point at infinity. Then a uniformizer at P is given by $y/x = 1/t$. We can compute that
$$\operatorname{ord}_{[0,1]}(dt) := \operatorname{ord}_{[0,1]}\left(\frac{dt}{d(1/t)}\right) = \operatorname{ord}_{[0,1]}\left(\frac{dt}{-t^{-2}dt}\right) = \operatorname{ord}_{[0,1]}(-t^2) = \operatorname{ord}_{[0,1]}(-x^2/y^2) = -2.$$

In conclusion, we have that the divisor of the differential $\omega = d(x/y)$ on C is

$$\operatorname{div} d(x/y) = -2 \cdot [1, 0].$$

(Lecture 7: February 3, 2021)

Example 2.7.8. Let $x, f \in \overline{K}(C)$ be nonzero, $P \in C$, and $x(P) = 0$. Consider the differential $\omega = f dx$. We will compute the order of vanishing of ω at P.

Let t be a uniformizer at P, so $\operatorname{ord}_P(t) = 1$. Let $n = \operatorname{ord}_P(x)$, so $x = ut^n$, with $\operatorname{ord}_P(u) = 0$. Then

$$dx = nut^{n-1}dt + t^n du = \left(nut^{n-1} + t^n \frac{du}{dt}\right) dt,$$

so by definition, the order of vanishing of ω at P is

$$\operatorname{ord}_P(\omega) := \operatorname{ord}_P\left(\frac{f dx}{dt}\right) = \operatorname{ord}_P\left(\frac{f\left(nut^{n-1} + t^n \frac{du}{dt}\right) dt}{dt}\right) = \operatorname{ord}_P(f) + \operatorname{ord}_P(t^{n-1}) + \operatorname{ord}_P\left(nu + t\frac{du}{dt}\right).$$

Recall that du/dt is regular at P because $\operatorname{ord}_P(u) = 0$. Thus, if we assume that $n \neq 0$, then this implies

$$\operatorname{ord}_P(f dx) = \operatorname{ord}_P(f) + \operatorname{ord}_P(x) - 1.$$

This is exactly what one should expect; when you differentiate something, the order that it vanishes at a point P should decrease by 1.

A small caveat: we assumed in this compuation $n \neq 0$, which is only the case if $\operatorname{char} K = 0$ or if $p \nmid \operatorname{ord}_P(x)$. In the case where $\operatorname{char} K > 0$ and $p \mid \operatorname{ord}_P(x)$, we must instead conclude that

$$\operatorname{ord}_P(f dx) \geq \operatorname{ord}_P(f) + \operatorname{ord}_P(x).$$

Definition 2.7.9. A differential $\omega \in \Omega_C$ is *holomorphic* if $\operatorname{ord}_P(\omega) \geq 0$ for all $P \in C$, and *nonvanishing* if $\operatorname{ord}_P(\omega) \leq 0$ for all $P \in C$.

Given nonzero differentials $\omega_1, \omega_2 \in \Omega_C$, we know that $\omega_1 = f\omega_2$ for some $f \in \overline{K}(C)$, as Ω_C is a 1-dimensional $\overline{K}(C)$-vector space. This implies that $\operatorname{div}(\omega_1) = \operatorname{div}(f) + \operatorname{div}(\omega_2)$. In other words, the divisors of any two nonzero differentials differ by the divisor of a some rational function. We write this as

$$\operatorname{div}(\omega_1) \sim \operatorname{div}(\omega_2).$$

This equivalence relation gives a well-defined divisor class, which turns out to be a very important invariant.

Definition 2.7.10. The *canonical divisor class* on C is the divisor class $[\mathrm{div}(\omega)] \in \mathrm{Pic}(C)$ corresponding to any nonzero differential $\omega \in \Omega_C$. We denote by K_C any canonical divisor, i.e. $K_C := \mathrm{div}(\omega)$ for any $\omega \in \Omega_C$.

Example 2.7.11. $\mathbb{P}^1$ has no holomorphic differentials (i.e. every differential form on $\mathbb{P}^1$ has at least one pole).

Proof. Consider the rational function $t := x/y \in \overline{K}(\mathbb{P}^1)$, and the corresponding nonzero differential dt. We computed above that $\mathrm{div}(dt) = -2 \cdot [1, 0]$. For any other nonzero differential ω, we know $\mathrm{div}(\omega)$ differs from $\mathrm{div}(dt)$ by the divisor of some rational function, say $\mathrm{div}(f)$. But any rational function $f \in \overline{K}(\mathbb{P}^1)$ has $\deg(f) = 0$, which implies that $\deg \mathrm{div}(\omega) = -2$. Thus, the differential ω has at least two poles. In summary, the degree of the divisor of a nonzero differential is an invariant, which we write as $\deg[K_C] = -2$. □

Example 2.7.12. Consider the projective cubic curve defined by the affine equation $y^2 = (x-\alpha)(x-\beta)(x-\gamma)$. Then one can compute that

$$\mathrm{div}(dx) = \mathrm{div}(y) = (\alpha) + (\beta) + (\gamma) - 3(\infty),$$

which implies that the differential dx/y has no poles or zeros. We'll come back to this example later.

2.8 Riemann-Roch

The goal now is to use the theory of divisors to specify poles and zeros of functions. For this, it's convenient to define:

Definition 2.8.1. A divisor $D = \sum n_P(P)$ is *positive* or *effective* if all $n_P \geq 0$, and we'll denote this by $D \geq 0$. We'll also write $D_1 \geq D_2$ to mean $D_1 - D_2 \geq 0$.

Example 2.8.2. Suppose we want to specify that "f has a pole at P of order at most m, and no other poles." Notationally, this is the same as specifying $\mathrm{div}(f) \geq -m(P)$.

Example 2.8.3. Suppose we want to specify that "f has a pole at P of order at most m, and no other poles, and it vanishes at Q to order at least k." Notationally, this is the same as specifying $\mathrm{div}(f) \geq -m(P)+k(Q)$.

Definition 2.8.4. For $D \in \mathrm{Div}(C)$, we define the set of functions

$$\mathcal{L}(D) := \left\{ f \in \overline{K}(C)^* : \mathrm{div}(f) + D \geq 0 \right\} \cup \{\infty\}.$$

In less classical language, this is really the sheaf of sections associated to the divisor D. Concretely, if $D = \sum n_P(P)$, then the condition $n_P \leq 0$ allows f to have a pole at P, and $n_P > 0$ forces f to have a zero at P.

Theorem 2.8.5. *(a) If* $\deg D < 0$*, then* $\mathcal{L}(D) = \{0\}$.

(b) $\mathcal{L}(D)$ *is a finite-dimensional* $\overline{K}$*-vector space.*

(c) If $D \sim D'$*, then* $\mathcal{L}(D) \cong \mathcal{L}(D')$ *(i.e. linearly equivalent divisors give isomorphic spaces of sections.) So*

$$\ell(D) := \dim_{\overline{K}} \mathcal{L}(D),$$

only depends on the divisor class $[D]$.

Proof. For (a), suppose $f \in \mathcal{L}(D)$, so $\mathrm{div}(f) + D \geq 0$. This means that $\deg \mathrm{div}(f) + \deg D \geq 0$. But we're assuming that $\deg D < 0$, which means that $\deg \mathrm{div}(f) > 0$, which is a contradiction, as nonzero rational functions have the same number of zeros and poles. For (b), the idea is to show that $\dim \mathcal{L}(D + (Q)) \leq \dim \mathcal{L}(D) + 1$. In words, each time you allow one more pole, you add at most one more dimension's worth of functions. □

A corollary of this proof is the upper bound $\ell(D) \leq \deg D$. But recall our goal: if we can show that $\mathcal{L}(D)$ has large dimension, then we know can force there to be some zeros at the points specified by D. So the upper bound $\ell(D) \leq \deg D$ is useless in this context. The goal is to show that $\ell(D) \geq \deg D$, perhaps minus

some sort of error term. This is in fact where the Riemann-Roch theorem comes from. But in order to state it properly, we need to discuss canonical divisors.

Let $K_C = \operatorname{div}(\omega)$ be a canonical divisor. Then $f \in \mathcal{L}(K_C)$ means $\operatorname{div}(f) + K_C \geq 0$, which means $\operatorname{div}(f) + \operatorname{div}(\omega) \geq 0$, which means $\operatorname{div}(f\omega) \geq 0$, which means $f\omega$ is holomorphic. One can actually exploit this correspondence to provide an isomorphism

$$\mathcal{L}(K_C) \cong \{\omega \in \Omega_C : \omega \text{ holomorphic}\}.$$

But remember that up to isomorphism $\mathcal{L}(K_C)$ only depends on the divisor class of K_C, which in turn only depends on C. Thus,

$$\ell(K_C) = \dim_{\overline{K}} \{\omega \in \Omega_C : \omega \text{ holomorphic}\}$$

is an integer invariant of C which that quantifies "how complicated" C is, since the bigger this integer is, the more differential forms there are on C.

Definition 2.8.6. The *genus* of the algebraic curve C is the dimension of the space of holomorphic differentials on C, i.e.,

$$g := \ell(K_C) = \dim_{\overline{K}} \{\omega \in \Omega_C : \omega \text{ holomorphic}\}.$$

Geometrically, one might prefer to define the genus to be the number of holes in the Riemann surface C. One can show that these are definitions are equivalent where they both make sense, for example over $\mathbb{C}$. But we want to work over an arbitrarily field, so the idea of obtaining the genus geometrically by traingulating and counting the number of holes makes no sense in this general setting.

Theorem 2.8.7 (Riemann-Roch)**.** *Let C be a smooth projective curve with genus g, and let K_C be a canonical divisor. Then for all divisors $D \in \operatorname{Div}(C)$,*

$$\ell(D) - \ell(K_C - D) = \deg D - g + 1. \tag{2.1}$$

Remark 2.8.8. Riemann originally proved $\ell(D) \geq \deg D - g + 1$, then he had his graduate student Roch provide the error term.

The point: the Riemann Roch theorem implies that $\ell(D) \geq \deg D - g + 1$. The RHS of this is a positive number so long as we take $\deg D$ large enough.

Corollary 2.8.9. *Let C/K be a smooth curve, K_C a canonical divisor. Then there exists an integer $g \geq 0$ such that for any divisor $D \in \operatorname{Div}(C)$,*

(a) $\ell(K_C) = g$ (i.e. our definition of g is correct)

(b) $\deg K_C = 2g - 2$.

(c) If $\deg D > 2g - 2$, then $\ell(D) = \deg D - g + 1$.

In other words (c) is saying that if our space of functions can have $2g-2$ poles, then the $\ell(K_C - D)$ term in (2.1) vanishes.

Proof. For (a), we'll apply Riemann-Roch with $D = 0$. This says that $\ell(0) - \ell(K_C) = \deg(0) - g + 1$. But $\ell(0)$ is the dimension of the space of functions with no poles, and the only functions that have no poles are the constants, so $\ell(0) = 1$. The zero divisor has degree 0, so this implies $\ell(K_C) = g$. For (b), we'll apply Riemann-Roch with $D = K_C$. This yields $\ell(K_C) - \ell(0) = \deg K_C - g + 1$. So by part (a), we're done. And for (c), if $\deg D > 2g-2$, then $\deg(K_C - D) = \deg K_C - \deg D < 0$ by (b). This implies that $\ell(K_C - D) = 0$, since you'd need to have a function which has more poles than it has zeroes, which is impossible. □

Example 2.8.10. Consider the case $C = \mathbb{P}^1$. We proved that $\mathbb{P}^1$ has no holomorphic differentials on $\mathbb{P}^1$ by arguing that $\deg \operatorname{div}(\omega) = -2$ for any nonzero $\omega \in \Omega_{\mathbb{P}^1}$. Therefore $\ell(K_C) = 0$, since this quantity is by definition the $\overline{K}$-dimension of holomorphic forms in Ω_C. But by the above corollary, $\ell(K_C) = g$. Therefore $g(\mathbb{P}^1) = 0$. This is consistent with our geometric notion of the genus, because $\mathbb{P}^1(\mathbb{C})$ is a sphere, so it has no holes.

Example 2.8.11. Still considering the case $C = \mathbb{P}^1$, part (c) of the corollary say that if for any divisor D with $\deg D \geq -1$, we know that $\ell(D) = \deg D + 1$. This tells us that there exist $\deg D + 1$ dimensions worth of holomorphic differentials on $\mathbb{P}^1$. This recovers an example that we already worked through; namely, we can construct holomorphic functions of the form

$$\prod_{i=1}^{d} \frac{\alpha_i X - \beta_i Y}{\gamma_i X - \delta_i Y}.$$

That is, they're just homogeneous functions with numerator and denominator of the same degree. The poles are where the denominator vanishes, and the zeros are where the numerator vanishes.

What can we say about the converse? If we assume that a curve has genus $g = 0$, then we claim that in fact $C \cong_{\overline{K}} \mathbb{P}^1$. This is not so clear a priori. This means that, for any curve C, if there are no nontrivial holomorphic differentials on C, then in fact the curve is $\mathbb{P}^1$. This result is evidence that there is value in studying the space of holomorphic differentials on a curve. The next most complicated question is: *what about the $g = 1$ case?* It turns out that these are the elliptic curves, and we'll turn to them in the next lecture.

2.9 Exercises

Exercise (Silverman example II.3.2). Let $e_1, e_2, e_3 \in \overline{K}$ be distinct, and consider the curve

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

Assume that $x^3 + Ax + B = (x - e_1)(x - e_2)(x - e_3)$, where all the e_i are distinct. We homogenize this to

$$\overline{C} : Y^2 Z = X^3 + AXZ^2 + BZ^3,$$

and denote the added point by $\mathcal{O}$. In particular, $x - e_i = \frac{X - e_i Z}{Z}$ and $y = Y/Z$ are rational functions on C, so should be able to compute their divisors. We will argue that

$$\operatorname{div}(x - e_i) = 2(P_i) - 2(\mathcal{O}), \qquad \operatorname{div}(y) = (P_1) + (P_2) + (P_3) - 3(\mathcal{O}). \tag{2.2}$$

Proof. In order to compute $\operatorname{ord}_{\mathcal{O}}(x - e_i)$, we will consider an affine patch that contains $\mathcal{O} = [0, 1, 0]$. A natural choice is the coordinates $(u, v) = (X/Y, Z/Y)$, and in this affine patch, C has coordinates

$$C : v = u^3 + Auv^2 + Bv^3. \tag{2.3}$$

We chose these coordinates because $\mathcal{O}$ has coordinates $(u, v) = (0, 0)$, thus

$$\operatorname{ord}_{\mathcal{O}}(x - e_i) = \operatorname{ord}_{(u,v)=(0,0)} \frac{\frac{X}{Y} - e_i \frac{Z}{Y}}{\frac{Z}{Y}} = \operatorname{ord}_{(u,v)=(0,0)} \frac{u - e_i v}{v}.$$

Let $m = \operatorname{ord}_{\mathcal{O}}(u)$ and $n = \operatorname{ord}_{\mathcal{O}}(v)$. Considering (2.3), we note that

$$\operatorname{ord}_{(u,v)=(0,0)}(u^3) = 3m \qquad \operatorname{ord}_{(u,v)=(0,0)}(Auv^2) = m + 2n \qquad \operatorname{ord}_{(u,v)=(0,0)}(Bv^3) = 3n,$$

so the RHS of (2.3) has $\mathcal{O}$-valuation $\geq \min\{3m, m + 2n, 3n\}$. As the LHS of (2.3) has $\mathcal{O}$-valuation n, we deduce that $n \geq \min\{3m, m + 2n, 3n\}$, hence $n \geq 3m$ since $m, n \in \mathbb{N}$. This implies that

$$\operatorname{ord}_{(u,v)=(0,0)}(u^3) = 3m \qquad \operatorname{ord}_{(u,v)=(0,0)}(Auv^2) \geq 4m \qquad \operatorname{ord}_{(u,v)=(0,0)}(Bv^3) \geq 9m,$$

so because the LHS and RHS of (2.3) must have the same $\mathcal{O}$-valuation, they both must have evaluation $n = 3m$. In summary, we've shown that $\operatorname{ord}_{\mathcal{O}}(v) = 3 \operatorname{ord}_{\mathcal{O}}(u)$.

In fact $\operatorname{ord}_{\mathcal{O}}(u) = 1$, and one can compute this directly from the definition of the local ring. Given this, we deduce that

$$\operatorname{ord}_{(u,v)=(0,0)} \frac{u - e_i v}{v} = \operatorname{ord}_{(u,v)=(0,0)}(u - e_i v) - \operatorname{ord}_{(u,v)=(0,0)}(v) = 1 - 3 = -2.$$

This implies that $x - e_i$ has a pole of order 2 at $\mathcal{O}$, and it clearly has no other poles. Since this is a rational function, $\deg(x - e_i) = 0$, which implies that it has two zeros. It certainly vanishes at $(e_i, 0)$ to order at least 1. Considering the valuation of $y^2 = (x - e_1)(x - e_2)(x - e_3)$ gives that $2\operatorname{ord}_{P_i}(y) = \operatorname{ord}_{P_i}(x - e_i)$, which implies that $x - e_i$ has a zero at P_i to order at least 2, as y has a zero at $(e_i, 0)$ to order at least 1. This implies that $\operatorname{div}(x - e_i) = 2(P_i) - 2(\mathcal{O})$, as needed.

We can do a similar computation for y,

$$\operatorname{ord}_{\mathcal{O}}(y) = \operatorname{ord}_{(u,v)=(0,0)}\left(\frac{Y}{Z}\right) = \operatorname{ord}_{(u,v)=(0,0)}\left(\frac{Y/Y}{Z/Y}\right) = \operatorname{ord}_{(u,v)=(0,0)}\left(\frac{1}{v}\right) = -3.$$

Clearly y has no other poles. As y is a rational function, $\deg(y) = 0$, so y has three zeros, and in fact it has a zero of degree at least 1 at each P_i because $2\operatorname{ord}_{P_i}(y) = \operatorname{ord}_{P_i}(x - e_i) = 2$. This completes the proof. □

Exercise (Silverman 2.2). Let $\phi : C_1 \to C_2$ be a nonconstant map of smooth curves, let $f \in \overline{K}(C_2)^*$, and let $P \in C_1$. Prove that

$$\operatorname{ord}_P(\phi^* f) = e_\phi(P)\operatorname{ord}_{\phi(P)}(f). \tag{2.4}$$

Proof. Let $t_{\phi(P)}$ be a uniformizer at $\phi(P)$, so we can write $f = t_{\phi(P)}^k g$ for some $g \in \overline{K}(C_2)$ which doesn't have a zero or pole at $\phi(P)$. Then (2.4) is equivalent to

$$\operatorname{ord}_P(t_{\phi(P)}^k \circ \phi) = e_\phi(P)\operatorname{ord}_{\phi(P)}(t_{\phi(P)}^k).$$

By definition, $e_\phi(P) = \operatorname{ord}_P(\phi^* t_{\phi(P)}) = \operatorname{ord}_P(t_{\phi(P)} \circ \phi)$, so we must show that

$$\operatorname{ord}_P(t_{\phi(P)}^k \circ \phi) = \operatorname{ord}_P(t_{\phi(P)} \circ \phi)\operatorname{ord}_{\phi(P)}(t_{\phi(P)}^k),$$

which is clear. □

Exercise (Silverman 2.3(a)(i)). Let $\phi : \mathbb{P}^1 \to \mathbb{P}^1$ be a rational map of degree $d \geq 1$. For $\alpha \in \mathbb{P}^1$, prove that

$$\sum_{\beta \in \phi^{-1}(\alpha)} e_\phi(\beta) = \deg \phi$$

for all $\alpha \in \mathbb{P}^1$.

Proof. Write $\phi = F(z)/G(z)$. Applying a linear change of variables if necessary, we may assume $\alpha \neq \infty$ and $\infty \notin \phi^{-1}(\alpha)$. List $\phi^{-1}(\alpha) = \{\beta_1, \ldots, \beta_r\}$, and notice that $F(z) - \alpha G(z) = 0$ exactly in $\{\beta_1, \ldots, \beta_r\}$. But because $F(z) - \alpha G(z)$ is a degree-d polynomial (since $\phi(\infty) \neq \alpha$) we can factor $F(z) - \alpha G(z) = c(z - \beta_1)^{e_1} \cdots (z - \beta_r)^{e_r}$. This implies that

$$\phi(z) - \alpha = (z - \beta_1)^{e_1} \cdot \frac{c(z - \beta_2)^{e_2} \cdots (z - \beta_r)^{e_r}}{G(z)},$$

where the factor on the right doesn't vanish or have a pole at β_1. This implies that $\operatorname{ord}_{\beta_i}(\phi(z) - \phi(\beta)) = e_i$, which in particular implies that $e_i = e_\phi(\beta_i)$. Therefore we can compute

$$\sum_{\beta \in \phi^{-1}(\alpha)} e_\beta(\phi) = \sum_{i=1}^{r} e_i = \deg \phi = d,$$

as needed. □

Exercise (Silverman 2.3(c)). Let $\phi : \mathbb{P}^1 \to \mathbb{P}^1$ be a rational map of degree d. Prove that[2]

$$2d - 2 = \sum_{\alpha \in \mathbb{P}^1} (e_\phi(\alpha) - 1).$$

[2]This is Hurwitz's theorem in the case of a nonconstant separable map $\phi : \mathbb{P}^1 \to \mathbb{P}^1$.

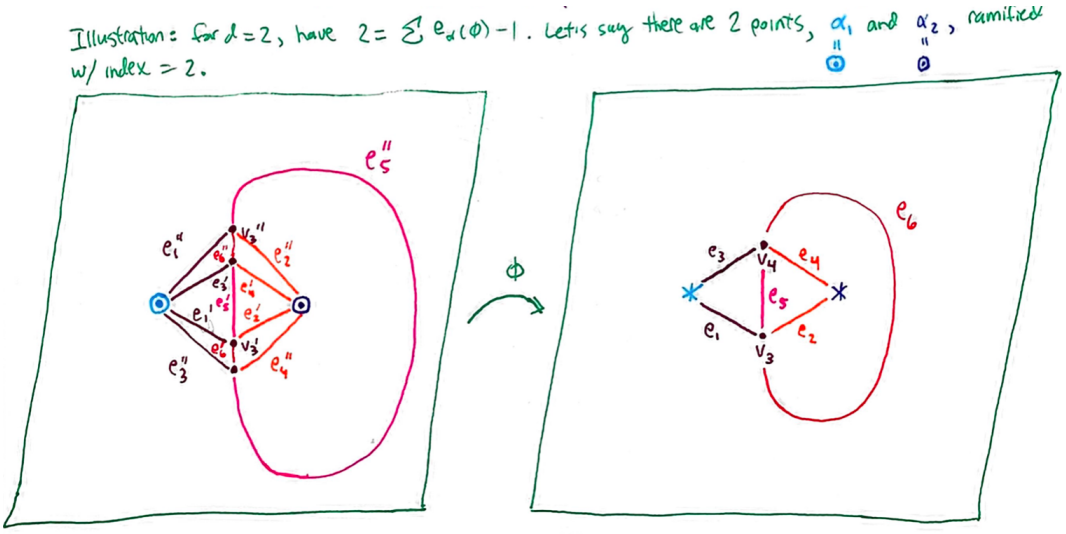
Proof. Take a sufficiently small triangulation of $\mathbb{C}$ with the property that the image of every ramification point of ϕ is a vertex, and let

$$\{v_i : 1 \le i \le V\}, \qquad \{e_i : 1 \le i \le E\}, \qquad \{f_i : 1 \le i \le F\}$$

be the vertices, edges, and faces of this triangulation. Observe that

$$\{\phi^{-1}(v_i) : 1 \le i \le V\}, \qquad \{\phi^{-1}(e_i) : 1 \le i \le E\}, \qquad \{\phi^{-1}(f_i) : 1 \le i \le F\}$$

is also a triangulation of $\mathbb{C}$, because ϕ is a local isomorphism away from ramification points. Say these sets have V', E' and F' elements, respectively.



This diagram illustrates phenomenon that, since ϕ is d-to-1 away from the ramification points, $E' = dE$ and $F' = dF$. And for a vertex v_i, by the previous exercise we know that $d = \sum_{\alpha \in \phi^{-1}(v)} e_\phi(\alpha)$. This implies that

$$d - \#\phi^{-1}(v) = \sum_{\alpha \in \phi^{-1}(v)} (e_\phi(\alpha) - 1).$$

Summing over all vertices v_i gives

$$dV - \sum_{i=1}^{V} \#\phi^{-1}(v_i) = \sum_{i=1}^{V} \sum_{\alpha \in \phi^{-1}(v_i)} (e_\phi(\alpha) - 1).$$

Next, we can extend this sum over all $\alpha \in \mathbb{P}^1$; this is justified because $e_\phi(\alpha) = 1$ for all α not in any $\phi^{-1}(v_i)$. This yields

$$dV = V' + \sum_{\alpha \in \mathbb{P}^1} (e_\phi(\alpha) - 1).$$

Finally, we can compute that

$$d(V - E + F) = dV - dE + dF = V' - E' + F' + \sum_{\alpha \in \mathbb{P}^1} (e_\phi(\alpha) - 1).$$

The result now follows from the fact that the Euler characteristic of the Riemann sphere is $\chi(\mathbb{P}^1) = V - E + F = V' - E' + F' = 2$. □

Exercise (Silverman 2.4). Let C be a smooth curve and let $D \in \mathrm{Div}(C)$. Without using the Riemann-Roch theorem, prove the following statements.

(a) $\mathcal{L}(D)$ is a $\overline{K}$-vector space.

(b) If $\deg D \geq 0$, then $\ell(D) \leq \deg D + 1$.

Proof. Recall that

$$\mathcal{L}(D) := \{f \in \overline{K}(C) : \operatorname{div}(f) \geq -D\}.$$

As $\operatorname{div}(0) = \infty \cdot [0]$ by convention, clearly $0 \in \mathcal{L}(D)$. If $f \in \mathcal{L}(D)$ and $c \in \overline{K}^*$, then $\operatorname{div}(cf) = \operatorname{div}(c) + \operatorname{div}(f) = 0 + \operatorname{div}(f) = \operatorname{div}(f)$, thus $cf \in \mathcal{L}(D)$. And if $f, g \in \mathcal{L}(D)$, then we must argue $f + g \in \mathcal{L}(D)$ as well. Writing $D = \sum_{P \in C} n_P [P]$, we'll argue in cases.

- *Case 1:* $P \in \operatorname{supp} D$ with $n_P < 0$. If we let t be a uniformizer for C at P, then we have $f = t^{n_f} u_f$ and $g = t^{n_g} u_g$ for $n_f, n_g > 0$ and $\operatorname{ord}_P(u_f), \operatorname{ord}_P(u_g) = 0$. Writing $m = \min\{n_f, n_g\}$, we compute $f + g = t^m (t^{n_f - m} u_f + t^{n_g - m} u_g)$, thus

  $$\operatorname{ord}_P(f+g) = \operatorname{ord}_P(t^m(t^{n_f-m}u_f + t^{n_g-m}u_g)) = \operatorname{ord}_P(t^m) + \operatorname{ord}_P(t^{n_f-m}u_f + t^{n_g-m}u_g) \geq m,$$

  which implies that $\operatorname{ord}_P(f+g) \geq \min\{\operatorname{ord}_P(f), \operatorname{ord}_P(g)\} \geq -n_P$, as needed.

- *Case 2:* $P \in \operatorname{supp} D$ with $n_P > 0$. We can take a uniformizer t of C at P and argue as in the previous case that $\operatorname{ord}_P(f+g) \geq -n_P$.

- *Case 3:* $P \notin \operatorname{supp} D$. In this case, as $\operatorname{div}(f), \operatorname{div}(g) \geq -D$, we know $\operatorname{ord}_P(f), \operatorname{ord}_P(g) \geq 0$, so f and g don't have a pole at P. We can take a uniformizer and compute as above that $\operatorname{ord}_P(f+g) \geq 0$.

This proves (a).

We'll prove (b) by induction on $\deg D$. The base case $\deg(D) = -1$ is clear, as we proved in Lecture 7 that $\ell(D) = 0$, hence $\ell(D) \leq \deg D + 1$. For the induction, it suffices to argue that for every $Q \in C$, we have

$$\ell(D + (Q)) \leq \ell(D) + 1.$$

Suppose $f, g \in \mathcal{L}(D + (Q)) - \mathcal{L}(D)$. If we define $m := -\operatorname{ord}_Q(D + (Q))$, then $\operatorname{ord}_Q(f) = \operatorname{ord}_Q(g) = m$, so if we let t be a uniformizer of C at P, then $f = t^m u$ and $g = t^m v$ for some $u, v \in \overline{K}(C)$ with $u(Q), v(Q) \neq 0$. Define $c := u(Q)/v(Q) \in \overline{K}^*$, then

$$\operatorname{ord}_Q(f - cg) = \operatorname{ord}_Q(t^m(u - cv)) = \operatorname{ord}_Q(t^m) + \operatorname{ord}_Q(u - cv) = m + \operatorname{ord}_Q(u - cv) \geq m + 1.$$

As $\mathcal{L}(D + (Q))$ is a vector space, the linear combination $f - cg$ is in $\mathcal{L}(D + (Q))$, but because

$$\operatorname{ord}_Q(f - cg) \geq -\operatorname{ord}_Q(D + (Q)) + 1 = -\operatorname{ord}_Q(D),$$

we deduce that $f - cg$ is in fact in $\mathcal{L}(D)$. This completes the proof. □

Exercise (Silverman 2.7). Let $F(X, Y, Z) \in K[X, Y, Z]$ be a homogeneous polynomial of degree $d \geq 1$, and assume that the curve C in $\mathbb{P}^2$ given by the equation $F = 0$ is nonsingular. Prove that

$$\operatorname{genus}(C) = \frac{(d-1)(d-2)}{2}. \tag{2.5}$$

Proof. We may assume $[0, 0, 1] \notin C$, by applying a linear change of variables if necessary. Therefore the map

$$\phi : C \to \mathbb{P}^1 : [X, Y, Z] \mapsto [X, Y]$$

is well-defined. So by the Hurwitz theorem, if we denote by g the genus of C, we have that

$$2g - 2 = -2 \cdot \deg \phi + \sum_{P \in C} (e_\phi(P) - 1). \tag{2.6}$$

It suffices to show the following two statements:

(a) $\deg \phi = d$.

(b) $\sum_{P \in C}(e_\phi(P) - 1) = d(d-1)$.

Given these facts, Hurwitz' theorem (2.6) implies (2.5).

Recall that $\deg \phi$ is the degree of the field extension $\phi^* : \overline{K}(\mathbb{P}^1) \to \overline{K}(C)$, where

$$\phi^*(\overline{K}(\mathbb{P}^1)) = \text{frac}\{\overline{f}(X,Y) : f \in \overline{K}[t]\}, \qquad \overline{K}(C) = \text{frac}\left(\overline{K}[X,Y,Z]/(F)\right).$$

Here $\overline{f} \in \overline{K}[X,Y]$ denotes the homogenization of $f \in \overline{K}[t]$. As f is a homogeneous polynomial of degree d, it follows that $[\overline{K}(C) : \phi^*(\overline{K}(\mathbb{P}^1))] = d$. For the second point, notice that for any $P \in C$, we have

$$F(P) = \frac{\partial F}{\partial y}(P) = 0$$

if and only if $P \in C$ with $e_\phi(P) > 1$. But by Bezout's theorem, the plane curves cut out by $F = 0$ and $\frac{\partial F}{\partial y} = 0$ intersect at exactly $d(d-1)$ points, $\deg(F) = d$ and $\deg(\frac{\partial F}{\partial y}) = d - 1$. □

Exercise (Silverman 2.8). Let $\phi : C_1 \to C_2$ be a nonconstant separable map of smooth curves.

(a) Prove that $\text{genus}(C_1) \geq \text{genus}(C_2)$.

(b) Prove that if C_1 and C_2 have the same genus g, then one of the following is true:

(i) $g = 0$.

(ii) $g = 1$ and ϕ is unramified.

(iii) $g \geq 2$ and ϕ is an isomorphism.

Proof. In this case, Hurwitz's theorem says that

$$2g_1 - 2 \geq (\deg \phi)(2g_2 - 2) + \sum_{P \in C_1}(e_\phi(P) - 1). \tag{2.7}$$

As $g_1, g_2 \geq 0$, and $\deg \phi \geq 1$, and $\sum_{P \in C_1}(e_\phi(P) - 1) \geq 0$, we can estimate $2g_1 - 2 \geq 2g_2 - 2$, so (a) is clear. For (b), let us assume C_1 and C_2 have the same genus $g > 0$. If $g = 1$, then (2.7) implies that $\sum_{P \in C_1}(e_\phi(P) - 1) \leq 0$, which is means ϕ must be unramified. And if $g = 2$, then (2.7) implies that $2g - 2 \geq (\deg \phi)(2g - 2)$, so it must be the case that $\deg \phi = 1$. □

(Lecture 8: February 5, 2021)

3 The geometry of elliptic curves

3.1 Ok, so... what actually *is* an elliptic curve?

Definition 3.1.1. An *elliptic curve* is a pair $(E, \mathcal{O})$ where E is a smooth projective curve of genus $g = 1$ and $\mathcal{O} \in E$. The elliptic curve E is *defined over* K if:

1. E as a projective algebraic set is defined over K.
2. $\mathcal{O} \in E(K)$.

3.2 Using Riemann-Roch to concretely classify elliptic curves

Our first task is to use the Riemann-Roch theorem to classify elliptic curves. As E is a projective variety, we ought to be able to embed $E \subseteq \mathbb{P}^N$ for some N. We will do this explicitly by finding functions $f_0, f_1, \ldots, f_N \in K(E)$, and we'll use these functions to create a rational map $[f_0, \ldots, f_N] : E \to \mathbb{P}^n$, which automatically extends to a morphism because E is a smooth curve. Crucially, the Riemann Roch theorem will let us construct these functions. But in order to apply Riemann-Roch, we'll need to choose good divisors on E. A natural collection of divisors to use is provided by $(\mathcal{O}), 2(\mathcal{O}), 3(\mathcal{O}), \ldots$. But we do these spaces of functions look like? By definition,

$$\begin{aligned}\mathcal{L}(n(\mathcal{O})) &= \{f \in K(E) : \operatorname{div} f + n(\mathcal{O}) \geq 0\} \\ &= \{f \in E : f \text{ has a pole at } \mathcal{O} \text{ of order at most } n, \text{ and no other poles}\}\end{aligned}$$

The Riemann-Roch theorem says that $\ell(D) + \ell(K_C - D) = \deg D - g + 1$ In our case, $g = 1$, so $\deg(K_C) = 2g - 2 = 0$. Therefore, if $\deg D > 0$, then $\ell(K_C - D) = 0$, so in this particular case the Riemann Roch theorem says that $\ell(D) = \deg D$. So now we can apply this with the divisors $D = n(\mathcal{O})$ with $n \geq 1$, which yields

$$\ell(n(\mathcal{O})) = \deg n(\mathcal{O}) = n.$$

Example 3.2.1. We'll provide a K-spanning set of $\mathcal{L}(n(\mathcal{O}))$ for small values of n.

- $\ell(\mathcal{O}) = 1$, thus $\mathcal{L}(\mathcal{O}) = \langle 1 \rangle$, since only the constant functions $E \to K$ have the right quantity of poles.
- $\ell(2(\mathcal{O})) = 2$, thus there exists some $x \in \overline{K}(E)$ with $\operatorname{ord}_{\mathcal{O}}(x) = -2$ such that $\mathcal{L}(2(\mathcal{O})) = \langle 1, x \rangle$.
- $\ell(3(\mathcal{O})) = 3$, thus there exists some $y \in \overline{K}(E)$ with $\operatorname{ord}_{\mathcal{O}}(y) = -3$ such that $\mathcal{L}(3(\mathcal{O})) = \langle 1, x, y \rangle$.
- $\ell(4(\mathcal{O})) = 4$, thus $\mathcal{L}(4(\mathcal{O})) = \langle 1, x, y, x^2 \rangle$ since $\operatorname{ord}_{\mathcal{O}}(x^2) = -4$. Crucially, note that x^2 is linearly independent from $\{1, x, y\}$ over K because it has a pole of higher order.
- $\ell(5(\mathcal{O})) = 5$, thus $\mathcal{L}(5(\mathcal{O})) = \langle 1, x, y, x^2, xy \rangle$ since $\operatorname{ord}_{\mathcal{O}}(xy) = -5$.
- $\ell(6(\mathcal{O})) = 6$, thus $\mathcal{L}(6(\mathcal{O})) = \langle 1, x, y, x^2, xy, x^3, y^2 \rangle$, but in fact x^3, y^2 are linearly dependent over this spanning set.

To recapitulate, there exist $x, y \in \overline{K}(E)$ with the properties that $\operatorname{ord}_{\mathcal{O}}(x) = -2, \operatorname{ord}_{\mathcal{O}}(y) = -3$, and x, y have no other poles. Additionally,

$$\langle 1, x, y, x^2, xy, x^3, y^2 \rangle \subseteq \mathcal{L}(6(\mathcal{O})).$$

In words, these 7 rational functions in $\overline{K}(E)$ each have a pole of order at most 6 at $\mathcal{O}$. As these are 7 living in a space of $\overline{K}$-dimension 6, there exists a nontrivial relation among these functions, which we'll write as

$$\alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 x^2 + \alpha_5 xy + \beta x^3 + \gamma y^2 = 0. \tag{3.1}$$

Observe that the rational functions $x, y \in \overline{K}(E)$ determine a nonconstant map

$$\phi : E \to \mathbb{P}^2 : P \mapsto [x(P), y(P), 1].$$

The image points in the affine patch $Z = 1$ are precisely the locus of points in $\mathbb{A}^2$ satisfying the equation (3.1). Furthermore, let us note that $\beta, \gamma \neq 0$, since as $\operatorname{ord}_{\mathcal{O}}(x^3) = \operatorname{ord}_{\mathcal{O}}(y^2) = -6$, so a dependence relation without one of these would in effect cancel a rational function of $\operatorname{ord}_{\mathcal{O}} = -6$ by ones with $\operatorname{ord}_{\mathcal{O}} > -6$.

Now we'll do some algebra to obtain a simpler equation from (3.1). If we let $x = -\beta\gamma X$ and $y = \beta^2\gamma Y$, then (3.1) has leading terms $-\beta^4\gamma^3 X^3 + \beta^4\gamma^3 Y^2$, so we can divide the resulting equation by $\beta^4\gamma^3$ (this is why we argued $\beta, \gamma \neq 0$.) In summary, there exist $x, y \in K(E)$ with $\operatorname{ord}_{\mathcal{O}} x = -2$ and $\operatorname{ord}_{\mathcal{O}} y = -3$ and x, y having no other poles, as well as constants $a_1, \ldots, a_6 \in K$, such that

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6. \tag{3.2}$$

This is how many books start by defining elliptic curves, as projective curves that have an affine equation of the above form; but this makes them look like they appear out of nowhere. In contrast, we saw that this equation arises because every smooth projective curve of genus 1 has an equation like this, thanks to the Riemann-Roch theorem. If $\operatorname{char} K \neq 2$ then we can complete the square in (3.2), writing this as

$$y^2 + a_1xy + a_3y = \left(y + \frac{1}{2}a_1x + \frac{1}{2}a_3\right)^2 - \left(\frac{1}{2}a_1x + \frac{1}{2}a_3\right)^2,$$

so if we let $Y := y + \frac{1}{2}a_1x + \frac{1}{2}a_3$ then we obtain

$$Y = x^3 + a_2x^2 + a_4x + a_6 - \frac{1}{4}(a_1x + a_3)^2.$$

Gathering terms and renaming variables, we get

$$Y^2 = X^3 + b_2X^2 + b_4X + b_6.$$

And if $\operatorname{char} K \neq 2, 3$, then we can write

$$X^3 + b_2X^2 + b_4X + b_6 = \left(x + \frac{1}{3}b_2\right) + (\text{something})x + (\text{something}).$$

In summary, we've shown the following.

Proposition 3.2.2. *Let $(E, \mathcal{O})$ be an elliptic curve. Then there exists $x, y \in K(E)$ with $\operatorname{ord}_{\mathcal{O}}(x) = -2$ and $\operatorname{ord}_{\mathcal{O}}(y) = -3$ and no other poles, as well as $A, B \in K$, and a surjective morphism*

$$\phi : E \to \left\{(X, Y) \in \mathbb{A}^2 : Y^2 = X^3 + AX + B\right\} \cup \{[0, 1, 0]\} : P \mapsto \begin{cases} (x(P), y(P)) & P \neq \mathcal{O} \\ [0, 1, 0] & P = \mathcal{O}. \end{cases}$$

We must strengthen that proposition. We do this as follows:

Theorem 3.2.3. *Let $(E, \mathcal{O})$ be an elliptic curve. Then there exists $x, y \in K(E)$ with $\operatorname{ord}_{\mathcal{O}}(x) = -2$ and $\operatorname{ord}_{\mathcal{O}}(y) = -3$ and no other poles, as well as $A, B \in K$, such that the map*

$$\phi : E \to \left\{[X, Y, Z] \in \mathbb{P}^2 : Y^2Z = X^3 + AXZ^2 + BZ^3\right\} \cup \{[0, 1, 0]\} : P \mapsto \begin{cases} [x(P), y(P), 1] & P \neq \mathcal{O} \\ [0, 1, 0] & P = \mathcal{O}. \end{cases}$$

is an isomorphism.

Proof. In steps:

(a) One must show that ϕ is bijective on points. We already know ϕ is surjective (as it's a nonconstant map between curves) but the injectivity is not as clear.

(b) One must show that $\operatorname{Image}(\phi)$ is smooth, because then we'll have a bijective map between smooth curves, so it's necessarily an isomorphism.

Both proofs use Riemann-Roch. We'll provide a proof of (a), and then a partial proof of (b). Suppose $\phi(P) = \phi(Q)$. If $P = Q = \mathcal{O}$, then $\phi(P) = \phi(Q) = [0, 1, 0]$. And if $P, Q \neq \mathcal{O}$, then $x(P) = x(Q) =: \alpha$ and $y(P) = y(Q) =: \beta$ since $\phi = [x, y, 1]$. Suppose towards a contradiction that $P \neq Q$. Let's look at $\operatorname{div}(x - \alpha)$. The rational function $x \in \overline{K}(C)$ has a pole of order 2 at $\mathcal{O}$ and no other poles, so subtracting the constant $\alpha \in \overline{K}$ from it doesn't change the poles; and if we plug in P or Q to $x - \alpha$ then we get zero, thus

$$\operatorname{div}(x - \alpha) = (P) + (Q) - 2(\mathcal{O}),$$

because $x - \alpha$ is a rational function on E implies it has the same quantity of zeros as poles. Next we look at $\operatorname{div}(y - \beta)$. The rational function $y - \beta$ has a triple pole at $\mathcal{O}$, and no other poles, and it must vanish at P and Q; since $\deg \operatorname{div}(y - \beta) = 0$, $y - \beta$ must have some other zero, say at $R \in E$. Then we have

$$\operatorname{div}(y - \beta) = (P) + (Q) + (R) - 3(\mathcal{O}).$$

Consider the rational function $(y-\beta)/(x-\alpha)$ on E. Then we can compute that

$$\operatorname{div}\left(\frac{y-\beta}{x-\alpha}\right) = ((P)+(Q)+(R)-3(\mathcal{O})) - ((P)+(Q)-2(\mathcal{O})) = (R)-(\mathcal{O}).$$

This implies that

$$\frac{y-\beta}{x-\alpha} \in \mathcal{L}((\mathcal{O})).$$

But $\mathcal{L}((\mathcal{O}))$ has dimension 1, and is in fact the set of constants. So therefore $y-\beta = c(x-\alpha)$ for some $c \in \overline{K}$. But $y-\beta$ has a pole of order 3 at $\mathcal{O}$, and $x-\alpha$ has a pole at order at most 2 at $\mathcal{O}$, which is a contradiction. This proves that ϕ is injective.

Now, we know the elliptic curve E corresponds bijectively to the affine curve $E : y^2 = x^3 + Ax + B$ union some point at ∞, because the homogenized projective curve $Y^2Z = X^3 + AXZ^2 + BZ^3$ intersects the line $Z = 0$ only at the point $[0,1,0] =: \infty$. We claim that in fact $[0,1,0]$ is a nonsingular point. To prove this, we dehomogenize by setting $Y = 1$ and obtain the polynomial $P(x,z) := z - (x^3 + Axz^2 + Bz^3)$. Then $\mathcal{O}$ corresponds to the point $(x,z) = (0,0)$, so we must show that $(0,0)$ is a nonsingular point on $V(P(x,y)) \subseteq \mathbb{A}^2$. But $\frac{\partial P}{\partial z}(0,0) = 1$, so the Jacobian at $(0,0)$ never vanishes, hence this is indeed nonsingular point. For completeness, one should check for non-singularity at the other points as well. □

In summary, we started with an elliptic curve $(E,\mathcal{O})$, which is just a smooth projective curve of genus $g = 1$. We showed that E can be modeled using an equation of the form $y^2 = x^3 + Ax + B$, which gives a nonsingular variety in $\mathbb{A}^2$. We can go the other way as well; namely, we can start with some equation $y^2 = x^3 + Ax + B$ and ask the following:

Question 3.2.4. What conditions on A and B ensure that $y^2 = x^3 + Ax + B$ has no singular points?

Solution. Write $f(x) = x^3 + Ax + B$ and $F(x,y) = y^2 - f(x) = 0$. Suppose $P = (\alpha,\beta) \in E$ is singular. This happens if and only if $F(\alpha,\beta) = 0$ and $\frac{\partial F}{\partial x}(\alpha,\beta) = \frac{\partial F}{\partial y}(\alpha,\beta) = 0$. By definition, this is a point (α,β) where $\beta^2 = f(\alpha)$ and $f'(\alpha) = 0$ and $2\beta = 0$. But this is true if and only if $\beta = 0$ and $f(\alpha) = 0$ and $f'(\alpha) = 0$, which is true if and only if (α,β) is of the form $(\alpha,0)$, where α is a repeated root of $f(x)$. In summary, we've shown that $y^2 = x^3 + Ax + B$ is nonsingular if and only if $x^3 + Ax + b$ has no repeated roots, i.e. if $f(x)$ is a separable polynomial. But a polynomial $f(x)$ is separable if and only if its discriminant is nonzero. In our case, $\operatorname{Disc}(x^3 + Ax + B) = -16(4A^3 + 27B^2)$. □

To recapitulate: if we start with an elliptic curve $(E,\mathcal{O})$, then we can find a model for that curve which is isomorphic to $y^2 = x^3 + Ax + B$ with $4A^3 - 27B^2 \neq 0$.

(Lecture 9: February 8, 2021)

Last time, we proved the following.

Theorem 3.2.5. *Let E/K be an elliptic curve. Then there exist $A, B \in K$ with $\Delta := 4A^3 + 27B^2 \neq 0$ such that E is isomorphic to*

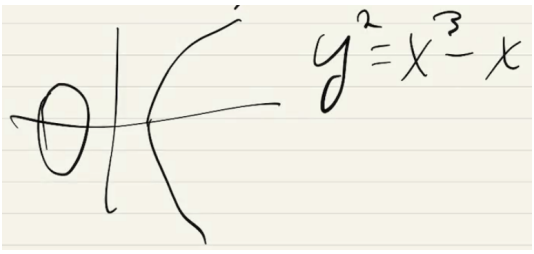
$$E : Y^2Z = X^3 + AXZ^2 + BZ^3,$$

with the base point $\mathcal{O}$ being mapped to $[0,1,0]$. This is called a Weierstrass equation *for E.*

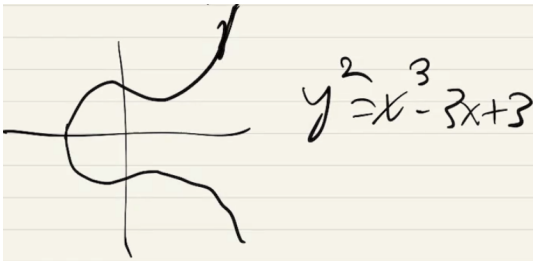
Note: we will always assume $\operatorname{char}(K) \neq 2, 3$. If $\operatorname{char} K = 2, 3$ then the equations end up messier.

3.3 Sample pictures of elliptic curves

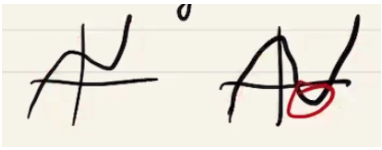
People usually illustrate an elliptic curve by drawing $E(\mathbb{R})$. One typical example:



Another example, which only has one real component:

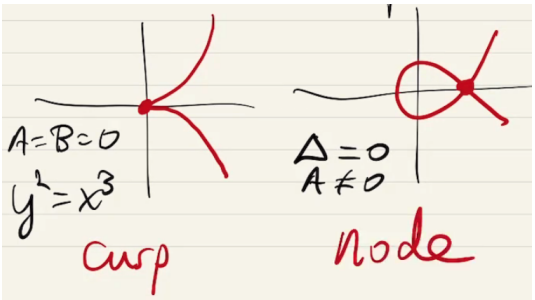


What determines whether there is one or two components? If we graph $y =$ the cubic, then there are two cases:



When we graph $y^2 =$ the cubic, it reflects the nonnegative values across the x-axis. So in the left case, there is only one component, whereas in the right case, the well below the x axis forces the elliptic curve to break into two components in $\mathbb{R}$.

We might consider an elliptic curve $E : y^2 = x^3 + Ax + B$ where $A, B \in \mathbb{Z}$ and $\Delta \neq 0$. If we reduce this modulo p, then we get $\tilde{E}/\mathbb{F}_p$, which is singular if $p \mid \Delta$. Over $\mathbb{R}$, singular curves have either a cusp or a node:



We'll see later that elliptic curves over any field generate a group if and only if the reduction is nonsingular; this is why we care about whether an elliptic curve is singularity.

3.4 When are two elliptic curves isomorphic?

Suppose an elliptic curve $(E, \mathcal{O})$ is isomorphic to curve cut out by both of the equations

$$y^2 = x^3 + Ax + B, \qquad (y')^2 = (x')^3 + A'x' + B'.$$

When does this happen? By construction, this means

$$\mathcal{L}(2(\mathcal{O})) = \langle 1, x \rangle = \langle 1, x' \rangle, \qquad \mathcal{L}(3(\mathcal{O})) = \langle 1, x, y \rangle = \langle 1, x', y' \rangle.$$

The former implies $x' = \alpha x + \beta$ for some $\alpha \neq 0$, and the latter implies $y' = \gamma y + \delta x + \epsilon$ for some $\gamma \neq 0$. Plugging this into the second equation for E gives

$$(y')^2 = (\alpha x + \beta)^2 + A'(\alpha x + \beta) + B' = \alpha^3 x^3 + 2\alpha^2 \beta x^2 + (\text{something})x + (\text{something}).$$

From this equation we can deduce that $\beta = 0$; similarly, by plugging $y' = \gamma y + \delta x + \epsilon$ into this equation we can deduce that $\delta = \epsilon = 0$. What this showed:

Proposition 3.4.1. *If $\{y^2 = x^3 + Ax + B\}$ is isomorphic to $\{(y')^2 = (x')^3 + A'x' + B'\}$, then $x' = \alpha x$ and $y' = \gamma y$ for some $\alpha, \gamma \neq 0$.*

Making the substitution $x' = \alpha x$ and $y' = \gamma y$ yields

$$y^2 = \gamma^{-2}\alpha^3 x^3 + (\text{something})A'x + (\text{something})B'.$$

But $y^2 - x^3$ cancels the pole of order 6 at $\mathcal{O}$ in the former equation for E, which implies that $y^2 - \gamma^{-2}\alpha^3 x$ must cancel the pole in the latter equation; this implies that $\gamma^2 = \alpha^3$. Let $u = \alpha/\gamma$; then we $y^2 = x^3 + u^4 A'x + u^6 B'$. Then comparing the two Weierstrass equations for E, we get $A = u^4 A'$ and $B = u^6 B'$. In summary, we have shown the following:

Theorem 3.4.2. *Consider the Weierstrass equations*

$$E : y^2 = x^3 + Ax + B, \qquad E' : y^2 : x^3 + A'x + B'.$$

Then E is isomorphic to E' over $\overline{K}$ if and only if there exists $u \in K^*$ such that $A' = u^4 A$ and $B' = u^6 B$.

We would like to use this theorem to develop a general procedure for finding out when two elliptic curves are isomorphic. The above theorem implies that, in the case where $B \neq 0$, E and E' are isomorphic if and only if

$$\frac{A^3}{B^2} = \frac{(A')^3}{(B')^2}.$$

So the idea is to utilize an equation similar to this one, but to put something in the denominator that isn't zero, because we want to allow the case where $B = 0$. Note that if $A' = u^4 A$ and $B' = u^B$, then the identity

$$\frac{A^3}{4A^3 + 27B^2} = \frac{(A')^3}{4(A')^3 + 27(B')^2}$$

holds. Conversely, if these two quantities are the same, you can deduce what u needs to be. In summary, the above quantity is an isomorphism invariant; it is the same regardless of which Weierstrass equation you're using to model an elliptic curve $(E, \mathcal{O})$. And this indeed solves the problem we faced earlier, as the denominator is nonzero whenever the elliptic curve is nonsingular.

Definition 3.4.3. Let $E : y^2 = x^3 + Ax + B$. The j-invariant of E is

$$j(E) := -1728\frac{(4A)^3}{\Delta},$$

where $\Delta := -16(4A^3 + 27B^2)$.

Here, $1728 = 12^3 = 2^6 3^2$, so as long as $\operatorname{char} K \neq 2, 3$, this is a well-defined invariant.

Theorem 3.4.4. *The j-invariant of E has the following properties:*

1. *$j(E)$ only depends on the isomorphism class of E.*
2. *E is isomorphic to E' over $\overline{K}$ if and only if $j(E) = j(E')$.*

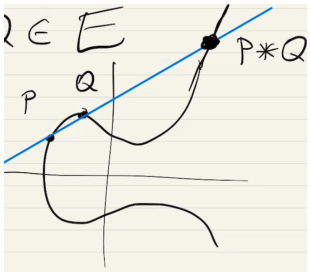
Remark 3.4.5. In fact, if you choose a different marked point $\mathcal{O}$, then you get the same j-invariant, but this takes more work to show.

Remark 3.4.6. This means is that genus 1 curves are paramereized by the affine line.

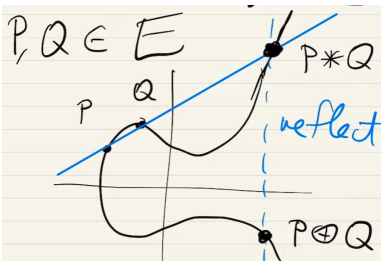
3.5 Addition on E, extrinsically

The miracle of elliptic curves is the group law. Namely, it turns out that $E = E(\overline{K})$ is an *algebraic group*, meaning there exists a group law on E which is given by rational functions of the coordinates. There are two ways to introduce this group law: via explicit equations, and using Riemann-Roch. Our goal is to understand this from both sides, but we'll start with the explicit equations.

How should we "add" point P and Q on E? The natural thing to do when given two points is to draw a line through them, and define the sum $P * Q$ to be the third point where that line intersects the curve:



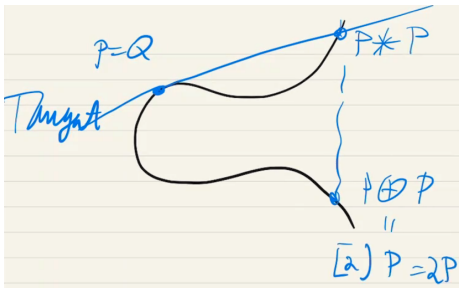
The problem with this approach is that this natural group law is garbage; for example, it doesn't have an identity. So we reflect $P * Q$ point across the x-axis, and call the resulting point $P + Q$.



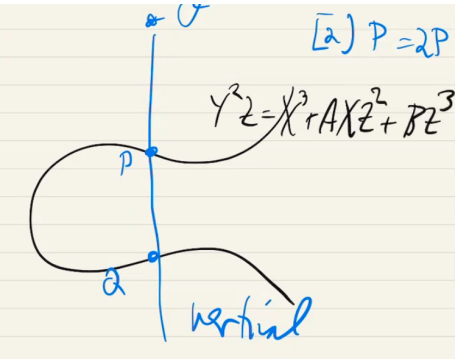
This turns out to give an adequate group law. But there are many technical points to take care of:

1. *Why does a line intersect the cubic in three points?* It's because a cubic intersecting a line, so generically it intersects in 3 points.

2. *How do we add a point to itself?* If we want the addition law to be continuous, then the natural thing to do is to take the tangent line at P:

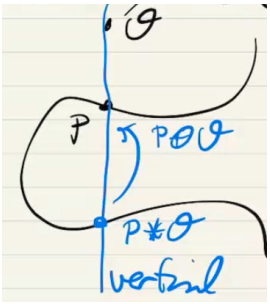


3. *How do we add two points that lie on the same vertical line?* There is no third intersection point on the affine plane, but if you take the projective coordinates, then you'll find the point at infinity is on this line.



Let's do that calculuation. Suppose P and Q are on the line $L : X = cZ$. If we plug this into $E : Y^2Z = X^3 + AXZ^2 + BZ^3$, then we get $Y^2Z = (c^3 + Ac + B)Z^3$. So either $Z = 0$, in which case $[X, Y, Z] = [0, 1, 0]$, or else if $Z = 1$ then we get the two points $[c, Y, 1]$ with $Y^2 = c^3 + Ac + B$; this gives the points P and Q. In summary, if P and Q are on a vertical line, then $P * Q$ is the point at ∞, so $P + Q = \mathcal{O}$.

4. *What is the identity operation?* By the above computation, $P + \mathcal{O} = P$, so $\mathcal{O}$ acts like the identity.



We summarize this extrinsic group law data in the following result.

Theorem 3.5.1. *Let $(E, \mathcal{O})$ be an elliptic curve with Weierstrass equation given by $y^2 = x^3 + Ax + B$. With the group law defined as above, the following are true:*

(a) $P + \mathcal{O} = P$

(b) Let P' be the reflection of P. Then $P + P' = \mathcal{O}$.

(c) The associative law $(P + Q) + R = P + (Q + R)$ holds.

We've proven the first two points; we'll come back to the associative law later, as it's tedious in this extrensic setting but more elegant in the intrinsic setting. What do we still have to do?

1. Find explicit formulas for $P + Q$.
2. Do addition more intrinsically (i.e. without referencing a specific Weierstrass equation).

The formulas are elementary to derive, but are messy.

Example 3.5.2. If $P = (x_1, y_1)$ and $Q = (x_2, y_2)$, then one can show that

$$x(P + Q) = \left(\frac{y_2 - y_1}{x_2 - x_2} - x_1 - x_2\right)$$

provided $x_1 \neq x_2$, and

$$x(2P) = \frac{x_1^4 - 2Bx_1^2 - 8Ax_1 - B^2}{4y_1^2}.$$

3.6 Addition on E, intrinsically

Now we'll define addition on E using the Riemann-Roch theorem. Given $P, Q \in E$, our task is to produce a third point in a natural way. Our corollary of the Riemann-Roch theorem says that if $\deg D \geq 1$, then $\ell(D) = \deg D$, because $g(E) = 1$. A natural divisor to consider is $(P) + (Q) - (\mathcal{O})$. This has degree 1, so $\mathcal{L}((P) + (Q) - (\mathcal{O}))$ has dimension 1, so it's of the form $\langle f_{P,Q} \rangle$, where $f_{P,Q} \in \overline{K}(E)$ is unique up to multiplication by $c \in \overline{K}^*$. In particular, $\operatorname{div}(f_{P,Q})$ depends only on P and Q. What does this divisor look like? By definition, $f_{P,Q} \in \mathcal{L}((P) + (Q) - (\mathcal{O}))$ means

$$\operatorname{div}(f_{P,Q}) + (P) + (Q) - (\mathcal{O}) \geq 0. \tag{3.3}$$

These are three cases to consider:

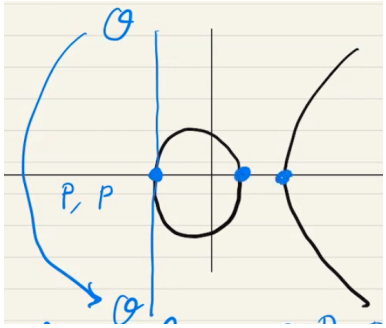
- If $f_{P,Q}$ has no poles, then it is a constant, which is a contradiction.
- If $f_{P,Q}$ has one pole, then the pole must be either P or Q by (3.3); since $\deg(f_{P,Q}) = 0$ and $f_{P,Q}$ must vanish at $\mathcal{O}$, it follows that $\operatorname{div}(f_{P,Q}) = (\mathcal{O}) - (P)$. This implies that we can think of the rational map $f_{P,Q} : E \to \overline{K}$ as a map into the Riemann sphere, $f_{P,Q} : E \to \mathbb{P}^1$. This map is one-to-one because it only has a single pole and a single zero, hence $f_{P,Q}$ is an isomorphism. This is a contradiction, as $g(E) = 1$ whereas $g(\mathbb{P}^1) = 0$.
- Thus by (3.3), $f_{P,Q}$ can have at most two poles, so it must have exactly two poles, which must be at P and Q. As $\deg(f_{P,Q}) = 0$, there exists a unique point $R \in E$ so that

  $$\operatorname{div}(f_{P,Q}) = -(P) - (Q) + (\mathcal{O}) + (R).$$

  We define this R to be the sum of P and Q.

(Lecture 10: February 10, 2021)

Example 3.6.1. If an elliptic curve in the affine plane has a vertical tangent line, then $P + P = \mathcal{O}$, which can be easily seen using the geometric definition of the group law.



In the elliptic curve pictured here, there are three vertical tangent lines. What does the Weierstrass equation $y^2 = x^3 + Ax + B$ tell us about these points of order two? By implicit differentiation,

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

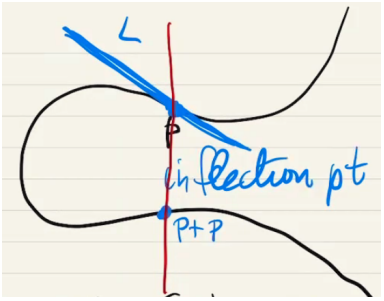
which is ∞ if and only if $y = 0$, which happens if and only if $x^3 + Ax + B = 0$. Say the roots of this cubic are α, β, and γ, then the points of order two are

$$E[2] = \{P \in E : 2P = \mathcal{O}\} = \{(\alpha, 0), (\beta, 0), (\gamma, 0), \mathcal{O}\}.$$

What are the abelian subgroups of order 4 whose elements all have order 2, except for the identity? The only option is

$$E[2] \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}.$$

Example 3.6.2. Suppose a line L hits a point P with multiplicity 3, so P is an inflection point of E. Then $L \cap E = \{P\} = \{P, P, P\}$, counted with multiplicity. Then $P + P = -P$, so $(P + P) + P = \mathcal{O}$ geometrically.



Then

$$E[3] = \{\text{inflection points}\} \cup \{\mathcal{O}\}$$

This is an abelian group of order 9 whose elements all have order 3, which implies that

$$E[3] \cong \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}.$$

Let's derive some explicit equations for this group law. Suppose we want to add $P_1 = (x_1, y_1)$ to $P_2 = (x_2, y_2)$ on $y = x^3 + Ax + B =: f(x)$. Let λ be the slope of the line through P_1 and P_2, or the slope of the tangent line if $P_1 = P_2$. Then we can compute

$$\lambda = \begin{cases} \frac{y_2 - y_1}{x_2 - x_1} & P_1 \neq P_2 \\ \\ \frac{3x_1^2 + A}{2y_1} & P_1 = P_2. \end{cases}$$

In this case, the line between P_1 and P_2 is given by

$$L : y = \lambda x + v, \qquad v := \frac{y_1 x_2 - y_2 x_1}{x_2 - x_1}.$$

We want to find the third intersection point, say $E \cap L = \{P_1, P_2, P_3\}$. Plugging the line into the equation for the curve, we have $(\lambda x + v)^2 = x^3 + Ax + B$, or equivalently, $x^3 - \lambda^2 x^2 + (A - 2\lambda - v)x + B - v^2$. This factors as $(x - x_1)(x - x_2)(x - x_3) = 0$. We can compare coefficients; the coefficient on x^2 is $-\lambda^2 = -x_1 - x_2 - x_3$, hence

$$(x(P_1 + P_2), y(P_1 + P_2)) = (\lambda^2 - x_1 - x_3, -\lambda x(P_1 + P_2) - \nu),$$

as the third point on the line has coordinates $(\lambda^2 - x_1 - x_3, \lambda x(P_1 + P_2) + \nu)$.

Example 3.6.3. Consider $E : y^2 = x^3 + 17$, This has points $P_1 = (-2, 3), P_2 = (-1, 4), P_3 = (2, 5)$. Then one can compute using these formulas that

$$P_2 + P_3 = \left(-\frac{8}{9}, -\frac{104}{107}\right), \qquad P_2 + P_2 = \left(\frac{127}{64}, -\frac{2561}{512}\right).$$

Remark 3.6.4. In the formula for $P_1 + P_2$, the sum $P_1 + P_2$ has rational entries if P_1 and P_2 do. This is not a priori true, as a line over $\mathbb{Q}$ intersecting a cubic over $\mathbb{Q}$ might have coordinates in some cubic extension. It's true in this case because if you factor a cubic equation, and find that two of the roots are rational, then the third must be rational as well.

Formalizing that observation:

Theorem 3.6.5. *Let E/K be an elliptic curve with Weierstrass equation $y^2 = x^3 + Ax + B$, with $A, B \in K$. Then $E(K) = \{(x, y) \in E : x, y \in K\} \cup \{\mathcal{O}\}$ is a subgroup of $E(\overline{K})$.*

Example 3.6.6. Let us return to $E : y^2 = x^3 + 17$. Then $E(\mathbb{Q})$ is a subgroup of $E(\mathbb{C})$, and in fact, it's a free group

$$E(\mathbb{Q}) = \mathbb{Z} \cdot (-2, 3) + \mathbb{Z} \cdot (2, 5).$$

This is a hard theorem to prove, and it's a consequence of the Mordell-Weil theorem, which we'll learn later.

Now, we return to the equivalent definition of the group law using divisors. Define a map

$$\phi : \mathrm{Div}^0(E) \to E : D \mapsto \phi(D),$$

where $\phi(D)$ is the unique point in E such that $D + (\mathcal{O}) \sim (\phi(D))$. Why does such a point exist? By the Riemann-Roch theorem, $\deg(D + (\mathcal{O})) = 1$ implies $\ell(D + (\mathcal{O})) = 1$, so $\mathcal{L}(D + \mathcal{O}) = \langle f \rangle$. Now, what is $\mathrm{div}(f)$? Because $\mathrm{div}(f) \geq -D - (\mathcal{O})$, and $\deg D = 0$, f must have zeroes at the holomorphic parts of $-D$; but in order to balance out the zeros forced upon f by $-D$, it must introduce poles, and it is only allowed to contain poles in the non-holomorphic part of $-D$, and at $\mathcal{O}$. But because $\deg(-D - (\mathcal{O})) = -1$, and because we need $\deg \mathrm{div}(f) = 0$, it must be that $\mathrm{div}(f) = -D - (\mathcal{O})+$ some extra point (which could potentially be $\mathcal{O}$) that we denote by $\phi(D)$. So by definition of linear equivalence, this implies $D + (\mathcal{O}) \sim (\phi(D))$.

Note that if $D \sim D'$, then by definition of ϕ, we have $(\phi(D)) \sim (\phi(D'))$. So there exists $g \in \overline{K}(E)$ with $\mathrm{div}(g) = (\phi(D)) - (\phi(D'))$. It follows that $g \in \mathcal{L}(\phi(D'))$. But the Riemann-Roch theorem tells us that $\dim \mathcal{L}(\phi(D')) = 1$. But $\mathcal{L}(\phi(D'))$ contains the constant functions, thus $g \in \overline{K}$ and $\phi(D) = \phi(D')$. In summary, ϕ descends to give a map on the quotient space, and we claim that it is in fact a bijection:

Theorem 3.6.7. *Let $(E, \mathcal{O})$ be an elliptic curve. Then the map*

$$\phi : \mathrm{Pic}^0(E) \to E : D \mapsto \phi(D)$$

is a bijection, with inverse given by

$$\psi : E \to \mathrm{Pic}^0(E) : P \mapsto [(P) - (\mathcal{O})].$$

Proof. We must argue that $\phi \circ \psi(P) = P$, and $\psi \circ \phi(D) \sim D$. For the first,

$$\begin{aligned} \phi \circ \psi(P) &= \phi([(P) - (\mathcal{O})]) \\ &= \text{the unique point } R \text{ so that } (P) - (\mathcal{O}) + (\mathcal{O}) \sim (R) \\ &= P. \end{aligned}$$

For the second, $D + (\mathcal{O}) \sim \phi(D)$ by definition of $\phi(D)$, so

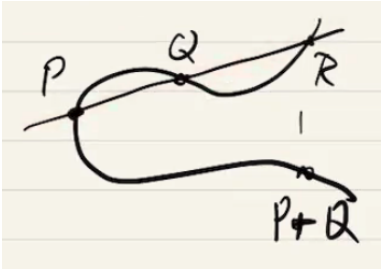
$$\psi(\phi([D])) = [(\phi(D)) - (\mathcal{O})] = [D].$$

□

But $\mathrm{Pic}^0(E)$ is an abelian group, so this bijection gives a group law on E. It would be nice if this gives the same group law as the one we defined using Bezout's theorem. We will prove that this is indeed the case.

Proposition 3.6.8. *The group law on E induced from the bijection $\psi : E \to \mathrm{Pic}^0(E)$ is isomorphic to the group law on E defined by intersecting lines on E.*

Proof. According to the group law coming from Bezout's theorem, for any $P, Q, R \in E$, we have that $P + Q + R = \mathcal{O}$ if and only if P, Q, and R are colinear.



As $P+Q=-R$ is equivalent to $P+Q+R=\mathcal{O}$, in order to show the group laws are the same, it suffices to show that P,Q,R are colinear if and only if $\psi(P)+\psi(Q)+\psi(R)=0$ in $\mathrm{Pic}^0(E)$. We will argue that

$$P,Q, \text{ and } R \text{ are colinear} \qquad \iff \qquad (P)+(Q)+(R) \sim 3(\mathcal{O}).$$

For $\implies$, we'll sketch the proof when P,Q, and R are distinct. Let $L: \alpha X+\beta Y+\gamma Z=0$ be the line which contains P,Q, and R, and consider the rational function

$$f=\frac{\alpha X+\beta Y+\gamma Z}{Z} \in \overline{K}(E).$$

Then f can only have poles when $Z=0$, so $\mathrm{div}_\infty(f)=n(\mathcal{O})$ for some $n\leq 0$. But the only zeros of f are where L intersects the curve, so f vanishes at least at the three points P,Q, and R; but by Bezout's theorem, L intersects E only at only three points, so $\mathrm{div}_0(f)=(P)+(Q)+(R)$ exactly. Because $\deg \mathrm{div}(f)=0$, we can deduce that $n=-3$, which implies that $(f)=(P)+(Q)+(R)-3(\mathcal{O})$, thus $(P)+(Q)+(R)$ is linearly equivalent to $3(\mathcal{O})$.

For $\impliedby$, suppose there exists $f\in\overline{K}(E)$ so that $(f)=(P)+(Q)+(R)-3(\mathcal{O})$. We must show that P,Q, and R are colinear. From the shape of $\mathrm{div}(f)$, it must be the case that

$$f=\frac{\text{some polynomial of } X,Y,Z}{Z^k}.$$

We can assume the numerator isn't divisible by Z, so in particular it doesn't vanish at $[0,1,0]$. This tells us that

$$\mathrm{ord}_{\mathcal{O}} f = -\mathrm{ord}_{Z=0} \frac{Z^k}{\text{the polynomial of } X,Y,Z} = -k\,\mathrm{ord}_{Z/Y=0}(Z/Y) = -3k,$$

because we computed in a previous lecture that the line $\{Z=0\}\cap E$ has a triple contact at $[0,1,0]$. But we know from the shape of $\mathrm{div}(f)$ that f has exactly three poles at $\mathcal{O}$, hence $k=1$. So by homogeneity, the numerator polynomial has to be a linear polynomial, so $f=\frac{\alpha X+\beta Y+\gamma Z}{Z}$, with P,Q, and R roots of the numerator because f vanishes at P,Q, and R. Therefore P,Q, and R are colinear, as we have explicitly constructed a line that they lie on. □

A useful fact: if $D\in\mathrm{Div}(E)$, and we write $D=\sum n_P(P)$, then $D\sim 0$ if and only if $\deg D=\sum n_P=0$. But this implies that $\sum n_P P=\mathcal{O}$m when we read this as addition in $\mathcal{O}$, as $\phi:\mathrm{Div}^0(E)\to E$ is a homomorphism, and $\deg\mathrm{div}(f)=0$. Recalling Riemann Roch, let K_C be the canonical divisor, so $K_C=\mathrm{div}(\omega)$ where ω is any nonzero differential form. Then $\mathcal{L}(K_C)$ is isomorphic to the space of holomorphic differential forms on C, and the dimension of this space $\ell(K_C)=g(C)$, and the degree of the canonical divisor is $\deg K_C=2g-2$ So, in our case, $g(E)=1$ implies $\ell(K_E)=1$, so up to scalar multiplication, there exists a unique $\omega_E\in\Omega_E$ with no poles, with $\deg\mathrm{div}(\omega_E)=0$. But this has no poles, which means it has no zeros.

In summary, Riemann-Roch implies there exists a differential form $\omega_E\in\Omega_E$ on E which is holomorphic, has no poles, is nonvanishing, and is unique up to a scalar. But we can find such a differential form explicitly:

Proposition 3.6.9. *Given $E: y^2=x^3+Ax+B$, we have that $dx/2y$ gives a nonvanishing holomorphic differential form.*

Proof. We'll prove part of the fact that $dx/2y$ is holomorphic. The only place this might fail to be holomorphic is at $\mathcal{O}$, as well as where $y=0$. We'll argue $y=0$ is impossible. By implicit differentiation, we can write $dx/2y$ as $dx/(3x^2+A)$; if we write $F(x):=x^3+Ax+B$, then this all implies that

$$\omega_E=\frac{dx}{2y}=\frac{dx}{F'(x)}=\frac{dx}{3x^2+A}.$$

But $2y=3x^2+A=0$ if and only if $F(x)=F'(x)=0$, which happens if and only if x is a repreated root of $F(x)$, which happens if and only if $(x,0)$ is a singular point of E. As E is nonsingular, it follows that $y\neq 0$ in the affine plane. And one can change coordinates to show that ∞ is also a nonsingular point. □

(Lecture 11: February 12, 2021)

Proposition 3.6.10. *If $P \in E$, then the translation map*

$$T_P : E \to E : Q \mapsto Q + P$$

is an automorphism of the genus 1*-curve* E.[3]

Proof. If $P = (x_1, y_1)$, then $T_P(x, y) = (x', y')$ where x', y' are rational functions in x, y, x_1, y_1. Thus $T_P : E \dashrightarrow E$ is a non-constant rational map of smooth curves, which implies it is a morphism. And of it's an *auto*morphism because its inverse is given by $T_P^{-1} = T_{-P}$. □

Theorem 3.6.11. *The addition law $\alpha : E \times E \to E : (P, Q) \mapsto P + Q$ is a morphism.*

Proof. We first argue that α is a rational map; so we must show that it's given by rational functions away from some Zariski-closed subset. If we assume (P, Q) satisfies $P, Q \neq \mathcal{O}$ and $P \pm Q \neq \mathcal{O}$ (i.e. we're throwing away some curves sitting on the surface $E \times E$) then $\alpha(P, Q) = (x', y')$, where x', y' are given by rational functions in x_P, y_P, x_Q, y_Q. But rational functions on smooth surfaces (unlike on smooth curves) can possibly have an indeterminacy locus, so we can't immediately conclude that the rational map $\alpha : E \times E \dashrightarrow E$ extends to a morphsim. To finish the proof, one can change coordinates and show that the excluded points are also given by rational maps.

Alternatively, given $\alpha : E \times E \dashrightarrow E$, we restrict to a morphism on the Zariski open subset, $\alpha : U \to E$. Fix $P_1, P_2 \in E$. Consider the chain of maps given by

$$E \times E \xrightarrow{T_{P_1} \times T_{P_1}} E \times E \overset{\alpha}{\dashrightarrow} E \xrightarrow{T_{-P_1 - P_2}} E,$$

and let $\phi : E \times E \dashrightarrow E$ be the composite map. We know $T_{P_1} \times T_{P_1}$ is a morphism because it's a morphism on each factor; and although $\alpha : E \times E \dashrightarrow E$ is a rational map, it's a morphism $\alpha : U \to E$; and $T_{-P_1 - P_2}$ is a morphism on E. An easy calculation reveals that the composition $\phi : E \times E \dashrightarrow E$ is given by $\phi(P, Q) = P + Q$, so it's a morphism on $(T_{-P_1} \times T_{-P_2})(U)$. Varying P_1 and P_2, we can cover $E \times E$. □

3.7 Isogenies

At this point, we've defined the objects of study, elliptic curves $(E, \mathcal{O})$. A natural next step of developing this theory is to present maps $(E_1, \mathcal{O}_1) \to (E_2, \mathcal{O}_2)$ that preserve the elliptic curve properties.

Definition 3.7.1. An *isogeny* from $(E_1, \mathcal{O}_1)$ to $(E_2, \mathcal{O}_2)$ is a non-constant morphism $\phi : E_1 \to E_2$ satisfying $\phi(\mathcal{O}_{E_1}) = \mathcal{O}_{E_2}$. We say E_1 and E_2 are *isogenous* if there exists a non-constant isogeny $E_1 \to E_2$.

Remark 3.7.2. One could equivalently define an isogeny as a non-constant rational map $E_1 \dashrightarrow E_2$ that maps base point to base point, because such a map is automatically a morphism.

Remark 3.7.3. Because non-constant maps between smooth curves are automatically surjective, it follows that isogenies are surjective.

Remark 3.7.4. Later we'll show the surprising fact that being isogenous is an equivalence relation, which isn't at all obvious. Furthermore, we'll show that an isogeny is necessarily also a group homomorphism; this too is not obvious, since a priori an isogeny is just a rational map.

If $\phi : E_1 \twoheadrightarrow E_2$ is an isogeny, then its *pullback* is

$$\phi^* : \overline{K}(E_2) \to \overline{K}(E_1) : f \mapsto \phi^* f = f \circ \phi,$$

and the *degree* of the isogeny is

$$\deg \phi := [\overline{K}(E_1) : \phi^* \overline{K}(E_2)].$$

We'll denote the *constant map* by

$$[0] : E_1 \to E_2 : P \mapsto \mathcal{O}_{E_2},$$

[3] Important to note: this says T_P is an automorphism of the projective variety E, but of course it's not an automorphism of the elliptic curve $(E, \mathcal{O})$, as T_P doesn't fix $\mathcal{O}$.

and we define $\deg[0] := 0$. Note this is necessarily a definition, as the corresponding function field extension would be a transcendental extension of K, so it'd be infinite. One can show that, given a chain of isogenies, the degrees multiply in towers, i.e.

$$E_1 \xrightarrow{\phi} E_2 \xrightarrow{\psi} E_3 \qquad \Longrightarrow \qquad \deg(\psi \circ \phi) = (\deg \psi)(\deg \phi).$$

Definition 3.7.5. The set of isogenies from E_1 to E_2 is denoted

$$\mathrm{Hom}(E_1, E_2) := \{\text{isogenies } E_1 \to E_2\} \cup \{[0]\}.$$

This is a group with the operation $(\phi_1 + \phi_2)(P) = \phi_1(P) + \phi_2(P)$. If $E = E_1 = E_2$, then the *endomorphism ring* of E is denoted

$$\mathrm{End}(E) := \mathrm{Hom}(E, E).$$

This is a group under the addition law, with multiplication given by composition.

It's not immediately obvious that the distributive law holds in $\mathrm{End}(E)$, one must check this carefully. We will also make use of the definition

$$\mathrm{End}_K(E) = \{\phi \in \mathrm{End}(E) : \phi \text{ defined over } K\}.$$

The most important isogeny:

Definition 3.7.6. Let $m \in \mathbb{Z}$. The *multiplication by m* isogeny, denoted $[m] : E \to E$, is defined as follows:

- For $m > 0$, $[m](P) := P + \cdots + P$, with m summands.
- For $m < 0$, $[m](P) := (-P) + \cdots + (-P)$, with m summands.
- For $m = 0$, $[0](P) := \mathcal{O}$.

Theorem 3.7.7. *If $m \neq 0$, then map $[m] : E \to E$ is an isogeny. In particular, for $m \neq 0$, $[m] : E \to E$ is not constant.*

Proof. We'll first argue that $[m]$ is a morphism. Note that $[2] : E \to E$ is an isogeny, we can write the doubling map as the composition of the morphisms

$$E \xrightarrow{P \mapsto (P,P)} E \times E \xrightarrow{\alpha} E.$$

Inductively, $[m+1](P)$ is given by the composition of morphisms

$$E \xrightarrow{P \mapsto (P,P)} E \times E \xrightarrow{[m] \times 1} E \times E \xrightarrow{\alpha} E,$$

which shows that $[m] : E \to E$ is a morphism for every $m > 0$. And of course the map $P \mapsto -P$ is a morphism, as it is simply the map that changes the y coordinate from y to $-y$.

It remains to show that $[m]$ is non-constant. Note that this is not immediate, because we don't have an explicit formula for multiplication by m. We'll sketch the proof for the case $\mathrm{char}\, K \neq 2$. The first step is to argue that $[2] \neq [0]$. The formula for twice a point is something like

$$[2](x, y) = \left(\frac{4x^4 - Bx^2 - 2Ax - B^2}{4(x^3 + Ax + B)}, \text{something else} \right).$$

This implies that $2[P] = \mathcal{O}$ if and only if $x^3 + Ax + B = 0$ or $P = \mathcal{O}$, hence

$$\{P \in E : [2](P) = \mathcal{O}\} = \{\mathcal{O}\} \cup \{\text{at most 3 other points}\}.$$

This implies that $[2] \neq [0]$, because the zero map sends every point on E to $\mathcal{O}$.

But in fact we can say more: because E is nonsingular, we know that $x^3 + Ax + B$ is separable, so

$$\#\{P \in E(\overline{K}) : [2]P = \mathcal{O}\} = 4.$$

So let $P_0 \in E(\overline{K})$ with $P_0 \neq \mathcal{O}$ but $[2]P_0 = \mathcal{O}$; if $m \in \mathbb{Z}$ is odd, then this implies $[m]P_0 = P_0 \neq \mathcal{O}$, which implies that $[m] \neq [0]$. Now, factor any $m \in \mathbb{Z}$ as $m = 2^k m_0$ for m_0 odd. Then $[m] = [2] \circ \cdots \circ [2] \circ [m_0]$ is a composition of nonconstant (hence surjective) morphisms, which implies that $[m]$ is surjective as well. $\square$

The space $\mathrm{Hom}(E_1, E_2)$ of isogenies $E_1 \to E_2$ naturally has the structure of an abelian group under addition, which means it's a $\mathbb{Z}$ module. In fact, we can say more:

Proposition 3.7.8. *If E_1 and E_2 are elliptic curves, then $\mathrm{Hom}(E_1, E_2)$ is a torsion-free $\mathbb{Z}$-module.*

Proof. Let $\phi \in \mathrm{Hom}(E_1, E_2)$, and suppose $[m] \circ \phi = [0]$. Then $\deg[m] \circ \phi = \deg[0] = 0$, thus $\deg[m] \cdot \deg \phi = 0$. As $[m]$ is a non-constant map, it follows that $\deg \phi = 0$, thus $\phi = [0]$. □

This endomorphism ring of E is a very important invariant.

Proposition 3.7.9. $\mathrm{End}(E)$ *is a ring of characteristic zero, with no zero divisors, but it is not necessarily commutative.*

Proof. The statement about characteristic follows from the previous proposition. For the statement about zero divisors, if $\phi \circ \psi = [0]$, then $\deg \phi \deg \psi = \deg[0] = 0$, so one of ϕ or ψ is the zero map. □

Remark 3.7.10. $\mathrm{End}(E)$ is only commutative if composition of endomorphisms is commutative. We'll show later that in characteristic zero, $\mathrm{End}(E)$ is indeed commutaitve. But this is not necessarily true in characteristic p; in this case, $\mathrm{End}(E)$ can actually be a quaternion algebra.

What we've just shown is that there is an injection

$$\mathbb{Z} \hookrightarrow \mathrm{End}(E) : m \mapsto [m].$$

By studying these multiplication by m maps, we'll get our hands on the harder arithmetic properties of the elliptic curve. A natural question is whether this map is onto, and the answer will be *sometimes yes and sometimes no.*

Definition 3.7.11. For $m \geq 1$, the *m-torsion subgroup* of E consisting of elements of order dividing m,

$$E[m] := \{P \in E : [m]P = \mathcal{O}\}.$$

More generally, the *torsion subgroup* of E is

$$E_{tors} := \bigcup_{n \geq 1} E[m] = \{P \in E : P \text{ has finite order}\}.$$

Example 3.7.12. We can explicitly compute $E[2]$, because

$$[2](x, y) = \left(\frac{something}{4(x^3 + Ax + B)}, something \right),$$

so $[2](x, y) = \mathcal{O}$ if and only if $x^3 + Ax + B = 0$, which is true if and only if $y = 0$. Thus

$$E[2] = \{\mathcal{O}\} \cup \{P : y_P = 0\}.$$

By nonsingularity, $x^3 + Ax + B$ has three distinct roots in the algebraic closure, so $E[2] = \{\mathcal{O}, Q_1, Q_2, Q_3\}$. Thus $E[2]$ is an abelian group of order 4 with exponent 2, so by the structure theorem for finitely generated abelian groups, it must be that

$$E[2] = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z},$$

which is the Klein 2-group. Important to highlight is that this is only if the characteristic is not 2; if the characteristic is 2, then we need to use the more complicated Weierstrass equation, and in that case, we have $E[2] = \{\mathcal{O}\}$ or $E[2] = \mathbb{Z}/2\mathbb{Z}$.

By considering the explicit duplication formula, it's easy to see that the finiteness of the degree implies that $\#E[m] < \infty$. In fact, one can say something much stronger. It'll take us a while to prove that the previous example is a shadow of a more general pattern:

Theorem 3.7.13. $E[m] \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$ *if* $\mathrm{char}\, K = 0$, *or if* $\mathrm{char}\, K = p$ *and* $p \nmid m$.

Definition 3.7.14. E has *complex multiplication* if $\mathrm{End}(E)$ is strictly bigger than $\mathbb{Z}$.

There are infinitely many curves over $\mathbb{Q}$ with complex multiplication, but they are much rarer.

Example 3.7.15. One can check directly that $E : y^2 = x^3 + x$ has an isogeny $\phi(x, y) = (-x, iy)$, where $i = \sqrt{-1}$. This actually gives us an isomorphism of rings

$$\mathbb{Z}[i] \hookrightarrow \mathrm{End}(E) : m + ni \mapsto [m] + [n] \circ \phi.$$

(Lecture 12: February 17, 2021)

We'll present an example of an isogeny of degree 2. The purpose of this example is to convince us we don't want to write down explcit equations for these things unless we really have to.

Example 3.7.16. Consider the elliptic curves

$$E : y^2 = x^3 + ax^2 + bx, \qquad E' : Y^2 = X^3 - 2aX^2 + (a^2 - 4b)X.$$

There are isogenies

$$\phi : E \to E' : (x, y) \mapsto \left(\frac{y^2}{x^2}, \frac{y(b - x^2)}{x^2}\right), \qquad \psi : E' \to E : (X, Y) \mapsto \left(\frac{Y^2}{4X^2}, \frac{Y(a^2 - 4b - X^2)}{X^2}\right),$$

which satisfy $\phi \circ \psi = [2]_{E'}$ and $\psi \circ \phi = [2]_E$.

Example 3.7.17. Consider an elliptic curve $E : y^2 = x^3 + Ax + B$, where $A, B \in K$ with $\mathrm{char}\, K = p \geq 5$. Then we can define the *Frobenius isogeny* as follows: for any prime power $q = p^r$, we can define

$$E^{(q)} : y^2 = x^3 + A^q x + B^q.$$

There exists a natural *q'th power Frobenius map*

$$\phi_q : E \to E^{(q)} : (x, y) \mapsto (x^q, y^q).$$

One can show that the j invariants are related by $j(E^{(q)}) = j(E)^q$, and the discriminants are related by $\Delta(E^{(q)}) = \Delta(E)^q$. Now suppose E is defined over $\mathbb{F}_q$, so $E^{(q)} = E$, and consider the points

$$\{P \in E(\overline{\mathbb{F}}_q) : \phi_q(P) = P\}.$$

These are the points $(x, y) \in E(\overline{\mathbb{F}}_q)$ so that $x^q = x$ and $y^q = y$, so $x, y \in \mathbb{F}_q$. So this set is just $E(\mathbb{F}_q)$. Later we'll be interested in counting how many points are in this set,

$$\#E(\mathbb{F}_q) = \#\{\text{fixed points of } \phi_q\},$$

or equivalently,

$$\#E(\mathbb{F}_q) = \#\{P \in E(\overline{\mathbb{F}}_q) : (1 - \phi_q)(P) = \mathcal{O}\}.$$

In summary, we've transformed the problem of counting points on varieties of finite fields into an algebraic problem, of counting the size of the kernel of $1 - \phi_q$.

We'll show now that any map on an elliptic curve given by rational functions is in fact a homomorphism. This is not at all obvious; note that lots of maps given by rational functions on the algebraic group $\mathbb{C}^*$ are not homomorphisms. This is only true in our case of elliptic curves because E is a compact algebraic group.

Theorem 3.7.18. *Suppose $\phi : E_1 \to E_2$ is an isogeny. Then ϕ is a homomorphism.*

Proof. Assume ϕ is nonzero, so it's surjective. We can look at the pushforward

$$\phi_* : \mathrm{Div}^0(E_1) \to \mathrm{Div}^0(E_2) : \sum_{P \in E_1} n_P(P) \mapsto \sum_{P \in E_1} n_P(\phi(P)).$$

Note that for any $f \in K(E_1)$, $\phi_*(\mathrm{div} f) = \mathrm{div}(\phi_* f)$; this means that ϕ_* maps principal divisors to principal divisors, so in particular, ϕ_* induces a well-defined map on the Picard groups,

$$\phi_* : \mathrm{Pic}^0(E_1) \to \mathrm{Pic}^0(E_2).$$

We consider the following diagram:

$$\begin{array}{ccc} \mathrm{Pic}^0(E_1) & \xrightarrow{\phi_*} & \mathrm{Pic}^0(E_2) \\ {\scriptstyle P\overset{\kappa_1}{\longmapsto}((P)-(\mathcal{O}))}\uparrow & & \uparrow{\scriptstyle Q\overset{\kappa_2}{\longmapsto}((Q)-(\mathcal{O}))} \\ E_1 & \xrightarrow{\phi} & E_2 \end{array}$$

The vertical maps are isomorphisms, and the diagram commutes, so $\phi = \kappa_2^{-1} \circ \phi_* \circ \kappa_1$ is a composition of homomorphisms. □

Ultimately, this proof relied on the isomorphism $E_1 \cong \mathrm{Pic}^0(E_1)$, which in turn relied in Riemann-Roch.

Corollary 3.7.19. *If $\phi : E_1 \to E_2$ is a nonzero isogeny, then $\ker\phi$ is a finite subgroup of E_1.*

Proof. Clearly $\ker\phi$ is a subgroup of E_1. As ϕ is a nonconstant map of smooth curves, we have the identity

$$\sum_{P\in\phi^{-1}(0)} e_\phi(P) = \deg\phi,$$

which implies that $\#\ker\phi \leq \deg\phi$. The next theorem shows that this actually an equality. □

Later we'll apply this important result to the case where E_1 and E_2 are the same curve, and $\phi = [m]$.

3.8 Isogenies, and Galois theory of elliptic function fields

An isogeny of elliptic curves $\phi : E_1 \twoheadrightarrow E_2$ induces an injection of the function fields

$$\phi^* : K(E_2) \hookrightarrow K(E_1) : f \mapsto f \circ \phi.$$

If ϕ is separable, then this is a Galois extension. We can say pretty precisely what it is:

Theorem 3.8.1. *If $\phi : E_1 \twoheadrightarrow E_2$ be an isogeny of elliptic curves, then the map*

$$\ker\phi \to \mathrm{Gal}(K(E_1)/\phi^*K(E_2)) : T \mapsto \tau_T^*$$

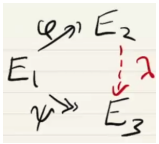
is an isomorphism, where $\tau_T : E_1 \to E_1$ is the translation by T map. In particular, $\#\ker\phi = \deg\phi$.

Part of proof. We will only verify that τ_T^* indeed fixes $\phi^*K(E_2)$. For any $P \in E_1$, we can compute that

$$\tau_T^*(\phi^* f)(P) = \tau_T^*(f\circ\phi)(P) = f\circ\phi\circ\tau_T(P) = f\circ\phi(T+P) = f\circ\phi(P) = \phi^* f(P),$$

as $T \in \ker\phi$. □

Corollary 3.8.2. *Suppose $\phi : E_1 \to E_2$ and $\psi : E_1 \to E_3$ are nonconstant separable isogenies. Assume that $\ker\phi \subseteq \ker\psi$. Then there exists a unique isogeny $\lambda : E_2 \to E_3$ such that the diagram*



commutes, i.e. $\lambda\circ\phi = \psi$.

Proof. By hypothesis, we know $\ker\phi \subseteq \ker\psi$. On the Galois level, the previous theorem implies there exists an inclusion

$$G(K(E_1)/\phi^*K(E_2)) \hookrightarrow G(K(E_1)/\psi^*K(E_3)).$$

So by Galois theory, there exists a unique inclusion of fields

$$\psi^*K(E_3) \subseteq \phi^*K(E_2) \subseteq K(E_1).$$

Recall that inclusions of functions fields of varieties give a rational map between the varieties; and since we're working with smooth curves, rational maps are automatically morphisms. Thus, there exists a unique morphism $\lambda : E_2 \to E_3$ so that

$$\phi^* \lambda^* K(E_3) = \psi^* K(E_3),$$

as $\phi^* \lambda^* = (\lambda\psi)^*$. Thus $\psi = \lambda \circ \phi$, and λ is an isogeny because $\lambda(\mathcal{O}) = \lambda \circ \phi(\mathcal{O}) = \psi(\mathcal{O}) = \mathcal{O}$. □

Now we turn to a fairly important construction. We've seen that an the isogeny out of E gives rise a finite subgroup of E, namely, $\ker \phi$. In fact the converse is true:

Theorem 3.8.3. *Let $\Phi \subseteq E$ be a finite subgroup. Then there exists a unique elliptic curve E' and a separable isogeny $\phi : E \to E'$ with $\ker \phi = \Phi$.*

Often we write $E' = E/\Phi$. Clearly E/Φ is a well-defined group; but what's to be shown is why E' defined this way is actually an elliptic curve (i.e., why must there be a Weierstrass equation model for the group E/ϕ) and why the projection map $E \to E/\Phi$ is a morphism of curves. There is a more general fact lurking here, namely, modding out any variety by a finite automorphism group gives a variety.

Proof. Each $\tau \in \Phi$ gives an automorphism $\tau_T : E \to E$, and its pullback is

$$\tau_T^* : K(E) \to K(E) : f \mapsto f \circ \tau_T,$$

which is an isomorphism because $(\tau_T^*)^{-1} = \tau_{-T}^*$. We can look at the subfield of $K(E)$ which is fixed by this automorphism,

$$K(E)^\Phi := \{f \in K(E) : \tau_T^* f = f (\forall T \in \Phi)\}.$$

Standard Galois theory tells us that $K(E)/K(E)^\Phi$ is Galois, and $G(K(E)/K(E)^\Phi) \cong \Phi$. In particular, this is a finite extension, so we have

$$K(E) \supseteq K(E)^\Phi \supseteq K.$$

The left extension finite and the whole tower has transendence degree 1, so $\operatorname{trdeg}_K K(E)^\Phi = 1$ implies $K(E)^\Phi = K(C)$ for some smooth algebraic curve C; further, there exists a map

$$\phi : E \to C \qquad \text{such that} \qquad \phi^* K(C) = K(E)^\Phi.$$

Let $P \in E, T \in \Phi$, and $f \in K(C)$. Then because $\phi^* f = f \circ \phi \in K(E)^\Phi$, we know τ_T^* fixes every element of $\phi^* K(C)$, hence

$$f \circ \phi(P + T) = f \circ \phi \circ \tau_T(P) = (\tau_T^* \circ \phi^*) f(P) = (\phi^* f)(P) = f \circ \phi(P).$$

But this identity holds for every $f \in K(C)$; as functions on C can distinguish points on C,[4] this actually implies that $\phi(P + T) = \phi(P)$ for all $P \in E$ and $T \in \Phi$. This implies that $\ker \phi = \Phi$.

It remains to show that C is an elliptic curve. Towards this, we claim that ϕ is unramified. For any $Q \in C$, we know that $\phi^{-1}(Q) \supseteq \{P + T : T \in \Phi\}$, so we can estimate

$$\#\Phi \leq \#\phi^{-1}(Q) = \sum_{P \in \phi^{-1}(Q)} 1 \leq \sum_{P \in \phi^{-1}(Q)} e_\phi(P) = \deg \phi = [K(E) : \phi^* K(C)] = [K(E) : K(E)^\Phi] = \#\Phi.$$

This implies that each $e_\phi(P) = 1$, so ϕ is indeed unramified. Therefore, by Riemann Hurwitz,

$$2g(C) - 2 = \deg \phi \cdot (2g(E) - 2) + \sum_{P \in E} (e_\phi(P) - 1).$$

But the RHS is zero, hence $g(C) = 1$. In summary, we've shown that $(C, \phi(\mathcal{O}))$ is an elliptic curve, and $\phi : E \to C$ is an isogeny with $\ker \phi = \Phi$. □

(Lecture 13: February 19, 2021)

[4] By Riemann Roch, I can find a function that has a pole at Q_1 and not at Q_2, for example.

3.9 Invariant differentials

Consider the elliptic curve given by $E : y^2 = x^3 + Ax + B$. Then the differential

$$\omega_E = \frac{dx}{2y} = \frac{dy}{3x^2 + A} \in \Omega_E$$

is holomorphic, i.e., $\operatorname{div}(\omega_E) = 0$. In fact, up to multiplication by a constant, this is the only holomorphic differential form, because $\dim \mathcal{L}(K_C) = 1$. The "invariant" here refers to the fact that this is differential form is invariant under translation, which we now prove:

Theorem 3.9.1. *If $\omega \in \mathcal{L}(K_E)$ is a generator, then*

$$\tau_Q^*(\omega) = \omega,$$

where $\tau_Q : E \to E : P \mapsto P + Q$ is the translation by Q map.

Proof. In principle, one can prove this with explicit formulas. The idea would be as follows: if we think of Q as fixed and P as the variable, then $x(P + Q)$ is some rational function of $x(P), y(P)$. Thus the differential

$$dx(P + Q) = \tau_Q^* dx(P) = (\text{ratl. fxn. of } x(P), y(P)) \cdot dx(P) + (\text{ratl. fxn. of } x(P), y(P)) \cdot dy(P).$$

But dy is some rational function times dx, so this whole thing is some rational function of $x(P), y(P)$ times $dx(P)$. All these rational functions can be computed explicitly; one just needs to verify that

$$\frac{dx(P + Q)}{2y(P + Q)} = \frac{dx}{2y}.$$

Alternatively, $\operatorname{div}(\tau_Q^*\omega) = \tau_Q^* \operatorname{div}(\omega) = \tau_Q^*(0) = 0$, thus $\tau_Q^*\omega$ is also a holomorphic differential on $E/$ Thus $\tau_Q^*\omega \in \mathcal{L}(K_E)$, which is a 1-dimensional K-vector space spanned by ω. This tells us that $\tau_Q^*\omega = c_Q \cdot \omega$ for some $c_Q \in K$. We claim that in fact $c_Q = 1$. As we argued abive, c_Q is some rational function of $x(Q), y(Q)$. Now for a clever idea: consider the map

$$E \to \mathbb{P}^1 : Q \mapsto c_Q.$$

This is a rational function. But E is a smooth curve, so this in fact this is a morphism, hence it's either onto or constant. But it's not onto, because it never hits ∞, as for every Q, this function gives a well-defined element $c_Q \in K$. So for all Q, we have $c_Q = c_{\mathcal{O}}$, which is 1. □

For motivation: calculus is basically a linearization tool. *We would like to use these differentials to linearize the addition law.*

Theorem 3.9.2. *Suppose we have isogenies $\phi, \psi : E' \to E$. Then the isogeny $\phi + \psi : E' \to E$ satisfies*[5]

$$(\phi + \psi)^*\omega = \phi^*\omega + \psi^*\omega.$$

Proof. Take two copies of E. Consider the following map given by addition on E,

$$\alpha : E \times E \to E : (x_1, y_1), (x_2, y_2) \mapsto (x_3, y_3),$$

so x_3, y_3 are some rational functions in x_1, y_1, x_2, y_2. Consider the invariant differential

$$\omega(x, y) = \frac{dx}{2y} = \frac{dy}{3x^2 + A}$$

on E. Using the chain rule, one can compute that

$$\omega(x_3, y_3) = f(x_1, y_1, x_2, y_2)\omega(x_1, y_1) + g(x_1, y_1, x_2, y_2)\omega(x_2, y_2)$$

[5]On the LHS, addition is done in $\operatorname{Hom}(E', E)$, which is defined using the group law on E; in contrast, on the RHS, addition is given in the vector space Ω_E. That's why this is such a valuable tool.

for some rational functions f and g that can be written out explicitly.

We claim that these rational functions f and g are actually constants equal to 1. One can check this directly using a computer algebra system. Alternatively, fix a point $Q \in E$, and set $x_2 = x(Q), y_2 = y(Q)$. Then the map α is just

$$P \mapsto \alpha(P, Q) = \tau_Q(P).$$

This implies that that $\omega(x_2, y_2) = 0$, as $\omega(x_2, y_2)$ can be written as a $K(x_1, y_2, x_2, y_2)$-linear combination of dx_2 and dy_2, and $dx_2 = d(\text{constant}) = 0$, and likewise $dy_2 = 0$. This implies that in the case where we're fixing $(x_2, y_2) = Q$, we have $\omega(x_3, y_3) = \tau_Q^* \omega(x_1, y_1)$. But we just proved that translation doesn't change the invariant differential, so this equation says $\tau_Q^* \omega(x_1, y_1) = \omega(x_1, y_1)$. This implies

$$\omega(x_3, y_3)|_{(x_2,y_2)=Q} = f(x_1, y_1, x(Q), y(Q))\omega(x_1, y_1) = \omega(x_1, y_1).$$

Thus, for all $Q \in E$, the map $E \to \mathbb{P}^1 : P \mapsto f(P, Q)$ is the constant map $P \mapsto 1$. In other words, $f(x_1, y_1, Q)$ is independent of x_1, y_1, which implies that $f(x_1, y_1, x_2, y_2) \in K(x_2, y_2)$. Also, $f(Q) = 1$ for all $Q \in E$. So in fact it doesn't depend on Q either. Thus $f = 1$ identically, and by symmetry $g = 1$ as well.

We've therefore proven that

$$\omega(x_3, y_3) = \omega(x_1, y_1) + \omega(x_2, x_2),$$

where (x_3, y_3) is the sum of the points (x_1, y_1) and (x_2, y_2). To finish the proof, consider the composition

$$E' \xrightarrow{\phi \times \psi} E \times E \xrightarrow{\alpha} E : (x', y') \mapsto (\phi(x', y'), \psi(x', y')) \mapsto (x, y) = \phi(x', y') + \psi(x', y').$$

Plugging this into the equation at the beginning of the paragraph, we get

$$\omega \circ (\phi + \psi)(x', y') = \omega \circ \phi(x', y') + \omega \circ \psi(x', y'),$$

or equivalently, $(\phi+\psi)^*\omega(x', y') = \phi^*\omega(x', y')+\psi^*\omega(x', y')$. This is exactly saying $(\phi+\psi)^*\omega = \phi^*\omega+\psi^*\omega$. □

What is the upshot of this highly technical result?

Corollary 3.9.3. *Let $m \in \mathbb{Z}$, and ω an invariant differential on E. Then*

$$[m]^*\omega = m \cdot \omega.$$

Proof. $[0]^*\omega = 0$ trivially, and $[1]^*\omega = \omega$ since this is the identity map, and by the previous proposition,

$$[m + 1]^*\omega = [m]^*\omega + [1]^*\omega = [m]^*\omega + \omega,$$

then induct. □

In summary, the differential takes the very complicated multiplication by m map on E, and turns it into multiplication by m on the vector space Ω_E.

Fact 3.9.4. If C is an algebraic curve and $0 \neq \omega \in \Omega_C$,[6] then $\phi : C' \to C$ is separable if and only if $\phi^*\omega \neq 0$.

Proof sketch. In characteristic 0, every extension is separable; in characteristic p, we get inseparable extensions when things are raised to the pth power. As $\phi^*\omega = \omega \circ \phi$, if ϕ raises things to p'th powers, then ω differentiates them and brings the powers down and kills it, which proves the forward implication. The converse says that having pth powers is the only way to kill things. □

Proposition 3.9.5. *Let $E/\mathbb{F}_q$, where $\mathbb{F}_q$ is a finite field of characteristic p. Let $\phi_q : E \to E : (x, y) \mapsto (x^q, y^q)$ be the q'th power Frobenius map. Let $m, n \in \mathbb{Z}$. Then the map*

$$m + n\phi_q : E \to E$$

is separable if and only if $p \nmid m$.

[6] Remember, Ω_C is a 1-dimensional $K(C)$-vector space.

Proof. Let ω be an invariant differential on E. Then $m + n\phi$ is separable if and only if $(m + n\phi)^*\omega \neq 0$ by the above fact; but this is equivalent to $[m]^*\omega + \phi^*[n]^*\omega \neq 0$ by the above theorem; but this is equivalent to $m\omega + n\phi^*\omega \neq 0$ by the above corollary. Note that ϕ^* is inseparable, since

$$\phi^*\omega = \phi^*\frac{dx}{2y} = \frac{d(x \circ \phi)}{2y \circ \phi} = \frac{d(x^q)}{2y^q} = \frac{qx^{q-1}dx}{2y^q}, = 0$$

as $q = 0$. Thus, the above is equivalent to $m\omega \neq 0$. But ω is a generator of the 1-dimensional $\mathbb{F}_q$ vector space of holomorphic differentials on E. The only integers which kill a nonzero element of the vector space are those which are zero in the field, i.e. those with $p \mid m$. This completes the proof. □

Remark 3.9.6. Why is separability useful? If ψ is separable, then

$$\#\psi^{-1}(Q) = \sum_{P \in \psi^{-1}(Q)} e_\psi(P) = \deg \psi$$

if ψ is unramified at Q. And for isogenies, we showed everything is unramified.

Proposition 3.9.7. *Define a map*

$$\mathrm{End}(E) \to K : \phi \mapsto a_\phi,$$

where a_ϕ is the constant such that $\phi^*\omega = a_\phi\omega$.

(a) This is a ring homomorphism.

(b) ϕ is in the kernel of this homomorphism if and only if ϕ is inseparable.

(c) If $\mathrm{char}\,(K) = 0$, then the homomorphism is injective, so in particular, $\mathrm{End}(E)$ is commutative.

(Lecture 14: February 22, 2021)

3.10 The dual isogeny

The basic idea: given an isogeny $\phi : E_1 \to E_2$, we want to create an isogeny $\hat{\phi} : E_2 \to E_1$ that is related in some interesting way to ϕ. A natural way to define an interesting homomorphism in the opposite direction is via the composition

$$E_2 \xrightarrow{\sim} \mathrm{Pic}^0(E_2) \xrightarrow{\phi^*} \mathrm{Pic}^0(E_1) \xrightarrow{\sim} E_1. \tag{3.4}$$

It turns out that the map we'll looking for will be this composition. But it's not clear at all why this map is given by rational functions, or in what quantifiable way it is related to ϕ. The intution here is that ϕ^* takes preimages, so it's essentially taking roots of polynomials, so the composition is indeed given by rational functions.

Theorem 3.10.1. *If $\phi : E_1 \to E_2$ is an isogeny, then:*

(a) There exists a unique isogeny $\hat{\phi} : E_2 \to E_1$ satisfying $\hat{\phi} \circ \phi = [m]$, where $m = \deg \phi$.

(b) $\hat{\phi}$ is equal to the composition (3.4).

Proof. For uniqueness, if $\hat{\phi}, \hat{\phi}'$ had this property, then $\hat{\phi} \circ \phi = \hat{\phi}' \circ \phi = [m]$, thus $(\hat{\phi} - \hat{\phi}') \circ \phi = [0]$. Thus, $\hat{\phi} - \hat{\phi}'$ is the constant map 0, since ϕ is nonconstant.

For existence, we claim that given a composition of isogenies

$$E_1 \xrightarrow{\phi} E_2 \xrightarrow{\psi} E_3,$$

if $\hat{\phi}$ and $\hat{\psi}$ exist, then $\widehat{\psi \circ \phi} = \hat{\phi} \circ \hat{\psi}$. To prove the claim, by uniqueness we just have to show the composition has the right property. Say $m = \deg \phi$ and $n = \deg \psi$. Then we must show

$$(\hat{\phi} \circ \hat{\psi}) \circ (\psi \circ \phi) = [mn].$$

But we can compute

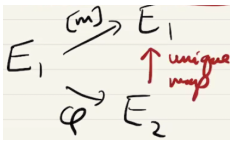
$$(\hat{\phi} \circ \hat{\psi}) \circ (\psi \circ \phi) = \hat{\phi} \circ \hat{\psi} \circ \psi \circ \phi = \hat{\phi} \circ [n] \circ \phi = [n] \circ \hat{\phi} \circ \phi = [n] \circ [m] = [mn].$$

Next, recall that any ϕ can be written in the form

$$\phi = (\text{separable map}) \circ (\text{Frobenius } q\text{-power map}).$$

Call the separable map ψ and Frobenius Φ_q. By what we just showed, it's enough to show that ψ has a dual, and Φ_q has a dual.

- *Case 1:* ϕ. Write $\deg \phi = m$. By separability, $\#\ker \phi = m$, so by Lagrange's theorem, $\ker \phi \subseteq \ker[m]$. But recall that if we have two different isogenies out of E_1 with the property that the kernel of one is contained in kernel of the other, then we can fill in the map so that the following diagram commutes:



  The map making this diagram commute is $\hat{\phi}$.

- *Case 2:* Φ_q. In this case, we're in characteristic p, and $q = p^r$. Then $\Phi_q = \Phi_p^r$, so it suffices find the dual isogeny of

$$\Phi_p : E \to E^{(p)}.$$

  Look at the multiplication by p map $[p] : E \to E$. The pullback of the invariant differential on E is $[p]^*\omega_E = p\omega_E$, as we proved in the previous lecture. But this is zero since we're in characteristic p. This implies that $[p]$ is inseparable, because the only way to kill the differential form is if we're raising all the variables to the p'th power, because then if we differentiate it we get zero. But any map decomposes as a separable map composed with a Frobenius map, so

$$[p] = (\text{separable map}) \circ \Phi_p^e.$$

  Then because $[p]$ is inseparable, this implies $e \geq 1$. Say λ is the separable map. Then we compute

$$(\lambda \circ \Phi_p^{e-1}) \circ \Phi_p = [p] = [\deg \Phi_p].$$

  Thus $\lambda \circ \Phi_p^{e-1} = \hat{\Phi}_p$ by definition, because when composed it with Frobenius, it is the multiplication by $\deg \Phi_p$ map.[7]

It remains to show that this dual actually agrees with the map given by the Picard groups; see the book for these details. □

Definition 3.10.2. Given an isogeny $\phi : E_1 \to E_2$, the *dual isogeny* is $\hat{\phi} : E_2 \to E_1$ with $\hat{\phi} \circ \phi = [m]_{E_1}$, where $m = \deg \phi$.

Proposition 3.10.3. *Some properties of the dual isogeny:*

1. $\phi \circ \hat{\phi} = [m]_{E_2}$.
2. *Given a composition of isogenies* $E_1 \xrightarrow{\phi} E_2 \xrightarrow{\lambda} E_3$, *we have* $\widehat{\lambda \circ \phi} = \hat{\phi} \circ \hat{\lambda}$.
3. *Given isogenies* $\phi, \psi : E_1 \to E_2$, *we have* $\widehat{\phi + \psi} = \hat{\phi} + \hat{\psi}$.
4. $\widehat{[m]} = [m]$.

[7] Note: the cases $e = 1, e = 2$ correspond to ordinary and supersingular elliptic curves.

Proof. For the first point, we compute

$$(\phi \circ \hat{\phi}) \circ \phi = \phi \circ (\hat{\phi} \circ \phi) = \phi \circ [m]_{E_1} = [m]_{E_2} \circ \phi,$$

so $(\phi \circ \hat{\phi} - [m]_{E_2})\phi = [0]$. But then $\phi \neq 0$ implies $\phi \circ \hat{\phi} - [m]_{E_2} = 0$. For the second point, we compute

$$(\hat{\phi} \circ \hat{\lambda}) \circ (\lambda \circ \phi) = \hat{\phi} \circ \hat{\lambda} \circ \lambda \circ \phi = \hat{\phi} \circ [n]_{E_2} \circ \phi = [n]_{E_1} \circ \hat{\phi} \circ \phi = [n]_{E_1} \circ [m]_{E_1} = [mn]_{E_1},$$

as needed. The third point is very time consuming to prove; see the textbook. For the fourth point, $\widehat{[0]} = 0$ and $\widehat{[1]} = 1$ obviously. Then by induction,

$$\widehat{[m+1]} = \widehat{[m] + [1]} = \widehat{[m]} + \widehat{[1]} = [m] + [1] = [m+1].$$

It's also easy to check that $\widehat{[-1]} = [-1]$, and then induct downwards. □

In particular, the fourth point allows us to compute the degree of the multiplication by m map.

Proposition 3.10.4. $\deg[m] = m^2$.

Proof. Let $d = \deg[m]$, and consider the map $[d]$. Then $[d] = \widehat{[m]} \circ [m]$ by definition of the dual isogeny. But $\widehat{[m]} = [m]$ by the fourth point, hence $[d] = [m] \circ [m] = [m^2]$. But recall that the map $\mathbb{Z} \to \mathrm{End}(E) : n \mapsto [n]$ is injective, thus $d = m^2$. □

A corollary of this:

Proposition 3.10.5. *Another property of the dual isogeny:*

5. $\deg \hat{\phi} = \deg \phi$.

Proof. We have $\hat{\phi} \circ \phi = [m]$ where $m = \deg \phi$, thus $\deg \hat{\phi} \cdot \deg \phi = \deg[m] = m^2$, so $\deg \hat{\phi} = m$. □

Proposition 3.10.6. *If $\deg \phi \neq 0$ in K, then ϕ is separable.*

Proof. Let $m = \deg \phi$. Then $m\omega_E = [m]^*\omega_E = \hat{\phi}^*\phi^*\omega_E$. Thus, in characteristic p, if $p \nmid m$, then $m\omega_E \neq 0$, which implies that $\phi^*\omega_E \neq 0$, which implies that ϕ is separable. □

This has an important consequence. Recall our notation $E[m] := \{P \in E : [m]P = \mathcal{O}\}$.

Corollary 3.10.7. *If $m \neq 0$ in K, then*

$$E[m] \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}.$$

Proof. In general $\#E[m] = \deg_s[m]$, which in this case is equal to $\deg[m]$ because $m \neq 0$ implies $\deg[m] = m^2 \neq 0$, thus $[m]$ is separable by what we showed above. But this implies that for all $d \mid m$, $\#E[d] = d^2$. One can show, using the structure theorem for finite abelian groups (the Smith normal form version) that this implies $E[m] = \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$. □

This turns out to be super important, as we'll study Galois groups by seeing how they act on these groups. To preview where we're going, write $G_K = \mathrm{Gal}(\overline{K}/K)$. Then G_K acts on the torsion points $E[m]$, as the multiplication by m map is defined over K so G_K commutes with $[m]$, as well as the fact that G_K respects addition on E. Thus, we get a map

$$G_K \to \mathrm{Aut}E[m],$$

where $\mathrm{Aut}E[m]$ denotes group automorphisms of $E[m]$. But $\mathrm{Aut}(\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}) \cong \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$, so in fact we get a 2-dimensional group representation

$$G_K \to \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z}),$$

We'll take m to be prime, in which case we'll get a representation of G_K into $\mathrm{GL}_2(\mathbb{F}_p)$; we'll also take $m = p^e$ and then take inverse limits and get a representation of G_K on $\mathbb{Z}_p$. These Galois representations are were the building blocks of Wiles' proof of the modularity theorem. How will we develop this theory? There is a natural pairing on 2-dimensional vector spaces, e.g., $\mathbb{R}^2 \wedge \mathbb{R}^2 \to \mathbb{R}$, given by the determinant. We'll want to do this intrindically (i.e. without choosing a basis) on $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$ as well, and we'll use the group law to define this pairing.

3.11 Exercises

Exercise (Silverman 3.4). Let $E/\mathbb{Q}$ be the elliptic curve

$$E : y^2 = x^3 + 17.$$

By inspection, this curve contains the points

$$P_1 = (-2, 3), \qquad P_3 = (2, 5).$$

Express each of the points $P_2 = (-1, 4), P_4 = (4, 9), P_5 = (8, 23), P_6 = (43, 282), P_7 = (52, 375)$, and $P_8 = (5234, 378661)$ in the form $[m]P_1 + [n]P_3$ with $m, n \in \mathbb{Z}$.

Solution. This is a straightforward computation using Sage. We include our code for completeness.

```
E = EllipticCurve([0,17])
P1 = E.point((-2,3))
P3 = E.point((2,5))

for m in range(-10,10):
        for n in range(-10,10):
                print((m*P1+n*P3))
```

We can now manually search through the output of this algorithm, yielding

$$\begin{aligned}
P_2 &= [-2]P_1 + [4]P_3 \\
P_4 &= [1]P_1 + [-1]P_3 \\
P_5 &= [-2]P_1 + [0]P_3 \\
P_6 &= [-1]P_1 + [2]P_3 \\
P_7 &= [3]P_1 + [-1]P_3 \\
P_8 &= [-4]P_1 + [3]P_3.
\end{aligned}$$

□

Exercise (Silverman 3.8). In two parts:

(a) Let $E/\mathbb{C}$ be an elliptic curve. One can show that there exist a lattice $L \subseteq \mathbb{C}$ and a complex analytic isomorphism of groups $\mathbb{C}/L \cong E(\mathbb{C})$. Assuming this fact, prove that

$$\deg[m] = m^2 \qquad \text{and} \qquad E[m] = \frac{\mathbb{Z}}{m\mathbb{Z}} \times \frac{\mathbb{Z}}{m\mathbb{Z}}.$$

(b) Let K be a field with $\operatorname{char}(K) = 0$ and let E/K be an elliptic curve. Use (a) to prove that $\deg[m] = m^2$.

Proof. Observe that $\mathbb{C}/L \cong \mathbb{C}/\mathbb{Z}^2$ as groups[8] so we may assume we're in this nicer case. By definition, $E[m]$ is the set of points with order dividing m, meaning it's the set of pairs $(x, y) \in \mathbb{C}/\mathbb{Z}^2$ with $(mx, my) \in \mathbb{Z}^2$, or equivalently, $(x, y) \in \mathbb{Z}^2/m$. There are exactly m^2 rational points in the fundamental domain with denominator dividing m, i.e., $\mathbb{Z}^2$-inequivalent points of the form $(a/m, b/m)$ with $a, b \in \mathbb{Z}$. Therefore $\#E[m] = m^2$. Concretely, the generators of $E[m]$ are given by

$$(1/m, 0), (0, i/m) \in \mathbb{C}/\mathbb{Z}^2.$$

Next, recall that $[m] : E \to E$ is an isogeny implies that $\deg[m] = \#\ker[m]$. But $\ker[m] = E[m]$ implies that $\deg[m] = m^2$. This proves part (a). For part (b), if E is defined over K, where $K \subseteq \mathbb{C}$, then E has a Weierstrass equation with coefficients in $\mathbb{C}$, so we can think of E as being defined over K. And if K can't be embedded into $\mathbb{C}$, then we can apply the Lefschetz principle. □

[8] Of course this isomorphism is not complex analytic, since it's not conformal.

Exercise (Silverman 3.9). Let E/K be an elliptic curve over a field K with $\mathrm{char}(K) \neq 2, 3$ and fix a homogeneous Weierstrass equation for E,

$$F(X_0, X_1, X_2) = X_1^2 X_2 - X_0^3 - AX_0X_2^2 - BX_2^3,$$

i.e., $x = X_0/X_2$ and $y = X_1/X_2$ are affine Weierstrass coordinates. Let $P \in E$.

(a) Prove that $[3]P = \mathcal{O}$ if and only if the tangent line to E at P intersects E only at P.

(b) Prove that $[3]P = \mathcal{O}$ if and only if the Hessian matrix

$$H := \left(\frac{\partial^2 F}{\partial X_i X_j}(P)\right)_{0 \leq i,j \leq 2}$$

has determinant 0. (This says that the the condition of the tangent line at P intersecting E only at P is equivalent to P being an inflection point of E.)

(c) Prove that $E[3]$ consists of nine points.

Proof. Part (a) follows directly from the geometric definition of the group law; the condition $[3]P = \mathcal{O}$ means $P + P = -P$, which means that the "third" point where the tangent line at P intersects E is $-(-P) = P$. By Bezout's theorem, the tangent line intersects E at exactly 3 points with multiplicity; thus the tangent line to E at P intersects E only at P. For part (b), we may assume $P \neq \mathcal{O}$; in this case, the point $P = (x_0, y_0)$ lies on the dehomogenized curve $E : y^2 = x^3 + Ax + B$. Consider the variable transformation $x = \overline{x} + x_0, y = \overline{y} + y_0$; then the point $(\overline{x}, \overline{y}) = (0, 0)$ corresponds to the point $(x, y) = (x_0, y_0)$. Plugging this into the dehomogenized Weierstrass equation for E gives the affinely transformed curve

$$\overline{E} : \overline{y}^2 + \alpha\overline{y} = \overline{x}^3 + \beta\overline{x}^2 + \gamma\overline{x}.$$

Note that there is no constant term because $(\overline{x}, \overline{y}) = (0, 0) \in \overline{E}$. If we implicitly differentiate the equation $\overline{y}^2 + \alpha\overline{y} - \overline{x}^3 - \beta\overline{x}^2 - \gamma\overline{x}$, then we get that $\frac{d\overline{x}}{d\overline{y}}(0, 0) = \gamma/\alpha$, so the tangent line through $(0, 0)$ is $\overline{y} = \frac{\gamma}{\alpha}\overline{x}$. This line is given parametrically by $\ell(t) = (\alpha t, \gamma t)$. Plugging this into the equation which cuts out $\overline{E}$, we obtain

$$0 = t^2(\alpha^3 t + \beta\alpha^2 - \gamma^2).$$

This equation implies that ℓ intersects E at $(0, 0)$ with multiplicity three if and only if $\beta\alpha^2 - \gamma^2 = 0$; by part (a), this implies that $[3]P = \mathcal{O}$ if and only if $\beta\alpha^2 - \gamma^2 = 0$. On the other hand, we can compute that the Hessian determinant of

$$\overline{F}(\overline{x}, \overline{y}, \overline{z}) = \overline{y}^2\overline{z} + \alpha\overline{y}\overline{z}^2 - \overline{x}^3 - \beta\overline{x}^2\overline{z} - \gamma\overline{x}\,\overline{z}^2$$

at $(0, 0, 1)$ is

$$\det \overline{H}|_P = \det \begin{pmatrix} \frac{\partial^2 F}{\partial \overline{x}^2} & \frac{\partial^2 F}{\partial \overline{x}\partial \overline{y}} & \frac{\partial^2 F}{\partial \overline{x}\partial \overline{z}} \\ \frac{\partial^2 F}{\partial \overline{x}\partial \overline{y}} & \frac{\partial^2 F}{\partial \overline{y}^2} & \frac{\partial^2 F}{\partial \overline{y}\partial \overline{z}} \\ \frac{\partial^2 F}{\partial \overline{x}\partial \overline{z}} & \frac{\partial^2 F}{\partial \overline{y}\partial \overline{z}} & \frac{\partial^2 F}{\partial \overline{z}^2} \end{pmatrix} |_{(0,0,1)} = \det \begin{pmatrix} -2\beta & 0 & -2\gamma \\ 0 & 2 & 2\alpha \\ -2\gamma & 2\alpha & 0 \end{pmatrix} = -8(\beta\alpha^2 - \gamma^2),$$

as needed. And for part (c), one can compute directly that the Hessian determinant has degree 3, and the equation which cuts out E has degree 3, so the intersection of the varieties E and $\{\det H = 0\}$ contains at most nine points. As $E[3]$ is an abelian group such that every point has order dividing three, by a divisibility argument we get that $\#E[3] = 9$, as needed. □

Exercise (Silverman 3.12). Let $m \geq 2$ be an integer, prime to $\mathrm{char}(K)$ if $\mathrm{char}(K) > 0$. Prove that the natural map

$$\mathrm{Aut}(E) \to \mathrm{Aut}(E[m])$$

is injective except for $m = 2$, where the kernel is $[\pm 1]$.[9]

[9] I credit this solution to Gal Porat on Mathematics Stack Exchange, link here.

Proof. Suppose $\sigma \in \mathrm{Aut}(E)$ satisfies $\sigma|_{E[m]} = \mathrm{id}|_{E[m]}$. Note that $\mathrm{id} - \sigma$ is an isogeny and $[m] : E \to E$ is a separable isogeny and with the property that $\ker[m] \subseteq \ker(\mathrm{id}-\sigma)$, because $[m]P = \mathcal{O}$ implies $\mathrm{id}(P)-\sigma(P) = P - P = \mathcal{O}$. Thus, there exists a unique isogeny $\lambda : E \to E$ making the following diagram commute,

$$\begin{array}{ccc} E & \xrightarrow{\ m\ } & E \\ & \searrow^{\mathrm{id}-\sigma} & \downarrow{\scriptstyle \lambda} \\ & & E. \end{array}$$

If we take the dual of the identity $\lambda \circ [m] = \mathrm{id} - \sigma$ then we obtain $[m] \circ \hat{\lambda} = \mathrm{id} - \sigma^{-1}$ because $[m]$ and id are self-dual, and because σ is an isomorphism implies its dual is its inverse. Composing these two equations yields

$$\begin{aligned} ([m] \circ \hat{\lambda}) \circ (\lambda \circ [m]) = (\mathrm{id} - \sigma^{-1}) \circ (\mathrm{id} - \sigma) &\implies [m^2 \deg \lambda] = [2] - \sigma - \sigma^{-1} \\ &\implies \sigma + \sigma^{-1} = [2 - m^2 \deg \lambda]. \end{aligned} \tag{3.5}$$

Now let us define the isogeny $\tau := 2\sigma - [2 - m^2 \deg \lambda]$. Then using (3.5), we can compute that

$$[\deg \tau] = \tau \circ \hat{\tau} = (2\sigma - [2 - m^2 \deg \lambda]) \cdot (2\sigma^{-1} - [2 - m^2 \deg \lambda]) = [4 - (2 - m^2 \deg \lambda)^2].$$

But $\deg \tau \geq 0$, which implies that $|2 - m^2 \deg \lambda| \leq 2$. Since $m \geq 2$ by hypothesis, this implies that $\deg \lambda \in \{0, 1\}$. We consider each case separately.

If $\deg \lambda = 0$, then $\lambda = 0$ implies $\mathrm{id} - \sigma = \lambda \circ [m] = 0$, so $\mathrm{id} = \sigma$. And if $\deg \lambda = 1$, then necessarily $m = 2$, so composing (3.5) with σ yields $\sigma^2 + \mathrm{id} = -2\sigma$, or equivalently, $(\sigma + \mathrm{id})^2 = 0$. Because $\mathrm{End}(E)$ is an integral domain, it follows that $\sigma = [-1]$. This completes the proof. □

Exercise (Silverman 3.21). Let $C/\overline{K}$ be a curve of genus one. For any point $O \in C$, we can associate to the elliptic curve (C, O) its j-invariant $j(C, O)$. In this exercise we will verify that the j-invariant is independent of the choice of base point O.

(a) Let (C, O) and (C', O') be curves of genus one with associated base points, and suppose that there is an isomorphism of curves $\phi : C \to C'$ satisfying $\phi(O) = O'$. Prove that $j(C, O) = j(C', O')$.

(b) Prove that given any two points $O, O' \in C$, there is an automorphism of C taking O to O'.

(c) Use (a) and (b) to conclude that $j(C, O) = j(C, O')$.

Proof. For (a), as (C, O) and (C', O') are curves of genus one, they have Weierstrass equations

$$C : y^2 = x^3 + Ax + B, \qquad C' : y^2 = x^3 + A'x + B',$$

where the point at infinity on each curve is the associated base point. We argued that these curves are isomorphic over $\overline{K}$ if and only if there exists $u \in K^*$ such that $A' = u^4 A$ and $B' = u^6 B$. On the one hand, we know that

$$j(C, O) = -1728 \frac{(4A)^3}{-16(4A^3 + 27B^2)},$$

and on the other hand, we can compute that

$$j(C', O') = -1728 \frac{(4A')^3}{-16(4(A')^3 + 27(B')^2)} = -1728 \frac{u^{12}(4A)^3}{-16(4u^{12}(A)^3 + 27u^{12}(B)^2)} = j(C, O).$$

For part (b), given two points $O, O' \in C$, consider the morphism

$$\tau_{O'-O} : C \to C : P \mapsto P + O' - O,$$

with addition given by the geometric group law. This translation map is an automorphism of C that satisfies $\tau_{O'-O} : O \mapsto O'$. And for part (c), we consider the isomorphism of based curves given by

$$\tau_{O'-O} : (C, O) \to (C, O').$$

By part (a), $j(C, O) = j(C, O')$, as needed. □

4 Elliptic curves over finite fields

(Lecture 15: February 24, 2021)

4.1 The Hasse bound

Today we'll discuss the degree map, and elliptic curves over finite fields.

Definition 4.1.1. An $\mathbb{R}$-valued *quadratic form* on an abelian group A is a function $d : A \to \mathbb{R}$ satisfying:

(a) The pairing $A \times A \to \mathbb{R} : (a, b) \mapsto d(a + b) - d(a) - d(b)$ is bilinear as a $\mathbb{Z}$-module homomorphism.

(b) $d(-a) = d(a)$.

Condition (a) makes d "quadratic" and condition (b) makes d a "form."

Proposition 4.1.2. *The map* $\deg : \mathrm{Hom}(E_1, E_2) \to \mathbb{Z}$ *is a positive definite quadratic form.*

Proof. Consider the map

$$\deg : \mathrm{Hom}(E_1, E_2) \times \mathrm{Hom}(E_1, E_2) \to \mathbb{Z} : (\phi, \psi) \mapsto \langle \phi, \psi \rangle := \deg(\phi + \psi) - \deg \phi - \deg \psi.$$

The quantity $\langle \phi, \psi \rangle$ is clearly an integer, so we can consider the multiplication by $\langle \phi, \psi \rangle$ map $[\langle \phi, \psi \rangle]$, and compute that

$$\begin{aligned}
[\langle \phi, \psi \rangle] &= [\deg(\phi + \psi)] - [\deg \phi] - [\deg \psi] \\
&= \widehat{(\phi + \psi)} \circ (\phi + \psi) - \hat{\phi} \circ \phi - \hat{\psi} \circ \psi \\
&= (\hat{\phi} + \hat{\psi}) \circ (\phi + \psi) - \hat{\phi}\phi - \hat{\psi}\psi \\
&= \hat{\phi}\psi + \hat{\psi}\phi.
\end{aligned}$$

This is indeed bilinear in ϕ and ψ. □

The positive-definiteness of the degree map can be used to prove a very deep theorem that was the start of a century's worth of research. Let $E/\mathbb{F}_q$ be an elliptic curve with $p \geq 5$ a prime and q a power of p. A very important and natural question:

How big is $\#E(\mathbb{F}_q)$*?*

In more elementary terms, if $q = p$, then we're simply asking a question in modular arithmetic; namely, we're asking how many solutions there are to $y^2 = x^3 + Ax + B \pmod{p}$. One can trivially bound the quantity of solutions as follows: the quantity of possible x values in $\mathbb{F}_q$ is q, and in a field, numbers have at most two square roots, thus

$$\#E(\mathbb{F}_q) \leq 2q + 1,$$

where the extra 1 comes from the point at infinity. But heuristically,

$$\#E(\mathbb{F}_q) \approx 1 + \sum_{x \in \mathbb{F}_q} 2 \cdot \mathrm{Prob}(x^3 + Ax + B \text{ is a square in } \mathbb{F}_q) \approx 1 + 2 \cdot \frac{q}{2} = q + 1.$$

We can formulate this exactly using the Legendre symbol:

$$\#E(\mathbb{F}_q) = 1 + \sum_{x \in \mathbb{F}_q} \left(\left(\frac{x^3 + Ax + B}{\mathbb{F}_q} \right) + 1 \right).$$

So our Heuristic says that $x^3 + Ax + B$ is about as likely to be a square as any given element of $\mathbb{F}_q$ is.

Another way of formulating this that is more useful:

$$E(\mathbb{F}_q) = \{P \in E(\overline{\mathbb{F}}_q) : \sigma(P) = P, \ \forall \sigma \in G(\overline{\mathbb{F}}_q / \mathbb{F}_q)\}.$$

But the Galois group $G(\overline{\mathbb{F}}_q/\mathbb{F}_q)$ is topologically generated by the q-th power Frobenius map Φ_q, which implies that

$$\begin{aligned} E(\mathbb{F}_q) &= \{P \in E(\overline{\mathbb{F}}_q) : \Phi_q(P) = P\} \\ &= \{P \in E(\overline{\mathbb{F}}_q) : (1-\Phi_q)(P) = \mathcal{O}\} \\ &= \ker(1-\Phi_q). \end{aligned}$$

In summary, we've translated the problem of counting $\mathbb{F}_q$-points on E to the problem of counting the kernel of an isogeny on E. We know that the size of kernel of an isogeny is the separable degree of the isogeny, but because $1-\Phi_q$ is separable, we can deduce that

$$\#E(\mathbb{F}_q) = \deg(1-\phi_q).$$

Now we've reduced the problem to estimating the degree of the difference of two isogenies. Abstractly, we have a ring, $\mathrm{End}(E)$, and positive-definieiy quadratic form on it, deg, and we want to estimate the value of the quadratic form on the difference of two elements in this endomorphism ring.

Lemma 4.1.3. *Suppose $d : A \to \mathbb{Z}$ is a positive definite quadratic form. Then*

$$d(\alpha+\beta) - d(\alpha) - d(\beta) \leq 2\sqrt{d(\alpha)d(\beta)}.$$

Proof. Let $L(\alpha,\beta) := d(\alpha+\beta) - d(\alpha) - d(\beta)$. Note that this is bilinear because d is a quadratic form. This means that for all $m, n \in \mathbb{Z}$, as d is positive-definite, we have

$$0 \leq d(m\alpha+n\beta) = L(m\alpha, n\beta) + d(m\alpha) + d(n\beta) = m^2 d(\alpha) + mnL(\alpha,\beta) + n^2 d(\beta).$$

But a quadratic form of this shape is positive definite if and only if its discriminant is non-positive, hence

$$L(\alpha,\beta)^2 - 4d(\alpha)d(\beta) \leq 0.$$

This implies that $|L(\alpha,\beta)| \leq 2\sqrt{d(\alpha)d(\beta)}$, as needed. □

Now, in the equation $\#E(\mathbb{F}_q) = \deg(1-\Phi_q)$, take $\alpha = 1$ and $\beta = -\Phi_q$; this lemma tells us that

$$|\deg(1-\Phi_q) - \deg(1) - \deg(-\Phi_q)| \leq 2\sqrt{\deg(1)\deg(-\Phi_q)},$$

which implies that

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

This proves a theorem conjectured by Artin in the 1920's and proven by Hasse in the 1930's:

Theorem 4.1.4 (Hasse)**.** *If $E/\mathbb{F}_q$ is an elliptic curve, then $|\#E(\mathbb{F}_q) - (q+1)| \leq 2\sqrt{q}$.*

Generalizations of this:

Theorem 4.1.5 (Weil)**.** *If $C/\mathbb{F}_q$ is a curve of genus g, then $|\#C(\mathbb{F}_q) - (q+1)| \leq 2g\sqrt{q}$.*

Theorem 4.1.6 (Deligne)**.** *If $V/\mathbb{F}_q$ is a smooth irreducible variety of dimension d, then*

$$|V(\mathbb{F}_{q^r}) - q^{dr}| \leq C_V q^{rd/2},$$

where the constant C depends only on the geometry of V, and in particular is independent of r.

It turns out that the Hasse bound is sharp for elliptic curves over $\mathbb{F}_p$, in the following sense:

Theorem 4.1.7. *Let $a \in \mathbb{Z}$ with $|a| < 2\sqrt{p}$. Then there exists an elliptic curve $E/\mathbb{F}_p$ with $\#E(\mathbb{F}_p) = p+1-a$.*

Another natural question: consider an elliptic curve $E : y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{Z}$. For each p, consider the elliptic curve $\tilde{E}_p/\mathbb{F}_p$ which is E reduced modulo p. Let $a_p = \#\tilde{E}_p(\mathbb{F}_p) - (p+1)$. Then Hasse's bound says $|a_p| \leq 2\sqrt{p}$, so a natural question to ask is: *how does a_p vary?* This was answered by Taylor and a bunch of other people.

Theorem 4.1.8 (Sato–Tate conjecture)**.** *Let $\theta_p \in [0,\pi]$ be such that $\cos\theta_p = a_p/2\sqrt{p}$. Then*

$$\lim_{T\to\infty} \frac{\#\{p \le T : \alpha \le \theta_p \le \beta\}}{\#\{p \le T\}} = \frac{1}{\sqrt{\pi}} \int_\alpha^\beta \sin^2 t dt.$$

In other words, the angles θ_p are distributed according to a $\sin^2$ distribution.

(Lecture 16: February 26, 2021)

4.2 The Tate module

Recall that

$$E[m] \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$$

for all M if $\operatorname{char} K = 0$, and for $\gcd(m, \operatorname{char} K) = 1$ if $\operatorname{char} K > 0$. This group of torsion points has extra structure because the Galois group acts on it. Specifically, if $\sigma \in G_K$ and $P \in E[m]$, then because the multiplication-by-m isogeny is defined over K, we can compute that

$$[m](P^\sigma) = ([m](P))^\sigma = \mathcal{O}^\sigma = \mathcal{O},$$

hence $P^\sigma \in E[m]$ as well. For the same reason, σ respects the group law, i.e. $(P+Q)^\sigma = P^\sigma + Q^\sigma$. Thus, $\sigma \in G_K$ induces a linear map $\sigma : E[m] \to E[m]$. This in turn induces a homomorphism

$$G_K \to \operatorname{Aut}(E[m]) \cong \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z}),$$

where the isomorphism $\operatorname{Aut}(E[m]) \cong \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ involves choosing a basis. This is a representation; if p is a prime, then this is called a *modular representation.* Our immediate goal is to fit these together to get a representation into GL_2 over something of characteristic zero. We'll do this by taking inverse limits.

We first recall a simpler case of this construction. If we fix a prime ℓ, then there are natural maps

$$\ell : \mathbb{Z}/\ell^{n+1}\mathbb{Z} \to \mathbb{Z}/\ell^n\mathbb{Z}$$

given by multiplication by ℓ, and the ℓ-adic integers are given by the inverse limit of these homomorphisms,

$$\mathbb{Z}_\ell := \varprojlim_n \mathbb{Z}/\ell^n\mathbb{Z}.$$

We can do an analogous construction using the torsion groups on an elliptic curve, starting with the natural multiplication-by-ℓ maps

$$\ell : E[\ell^{n+1}] \to E[\ell^n].$$

Definition 4.2.1. The *ℓ-adic Tate module is*

$$T_\ell(E) = \varprojlim_n E[\ell^n].$$

As each $E[\ell^n]$ is a $\mathbb{Z}/\ell^n\mathbb{Z}$ module, the Tate module has a natural structure as a $\mathbb{Z}_\ell$ module. As a group this is isomorphic to

$$T_\ell(E) \cong \varprojlim_n (\mathbb{Z}/\ell^n\mathbb{Z} \times \mathbb{Z}/\ell^n\mathbb{Z}) \cong \mathbb{Z}_\ell \times \mathbb{Z}_\ell.$$

But it is important to remember that this has a $\mathbb{Z}_\ell$ module structure, not just a $\mathbb{Z}$-module structure.

Definition 4.2.2. The *ℓ-adic representation of E/K* is

$$\rho_\ell = \rho_{\ell,E/K} : G_K \to \operatorname{Aut}(T_\ell(E)),$$

where $\sigma \in G_K$ acts on $T_\ell(E)$ by transforming the coherent sequences according to

$$(\dots, P_3, P_2, P_1) \overset{\sigma}{\mapsto} (\dots, P_3^\sigma, P_2^\sigma, P_1^\sigma).$$

If we choose a basis, then we get isomorphisms

$$\mathrm{Aut}(T_\ell(E)) \cong \mathrm{Aut}(\mathbb{Z}_\ell^2) \cong \mathrm{GL}_2(\mathbb{Z}_\ell) \subseteq \mathrm{GL}_2(\mathbb{Q}_\ell).$$

So in this way we get a representation into a characteristic zero field,

$$\rho_\ell : G_K \to \mathrm{GL}_2(\mathbb{Q}_\ell).$$

We can also obtain this representation by extending scalars,

$$G_K \xrightarrow{\rho_\ell} \mathrm{GL}(T_\ell(E)) \hookrightarrow \mathrm{GL}(T_\ell(E)) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell.$$

We will now pause briefly to motivate this construction. We'll consider an analogue of this construction in a more concrete case, namely the multiplicative group $G_m \cong \overline{K}^*$. In this group, consider the multiplication by m map

$$[m] : \overline{K}^* \to \overline{K}^* : z \mapsto z^m.$$

In this group, we have that

$$\ker[m] = \{\zeta \in \overline{K}^* : \zeta^m = 1\} =: \mu_m.$$

(This tells us that in the elliptic curve case, the torsion points of order m should really be thought of as m'th roots of unity, just under a different group.) We have the maps

$$\mu_{\ell^{n+1}} \xrightarrow{\zeta \mapsto \zeta^\ell} \mu_{\ell^n},$$

so we can define the Tate module in this case to be

$$T_\ell(G_m) := T_\ell(\mu) := \varprojlim_n \mu_{\ell_n}.$$

This gives an ℓ-adic representation

$$G_K \to \mathrm{Aut}(T_\ell(\mu)),$$

which is determined by the action of Galois acts on ℓ^n'th roots of unity. But abstractly, the target space is

$$\mathrm{Aut}(T_\ell(\mu)) \cong \mathrm{Aut}(\mathbb{Z}_\ell) = \mathrm{GL}_1(\mathbb{Z}_\ell) = \mathbb{Z}_\ell^*,$$

so we've actually constructed a map

$$G_K \to \mathbb{Q}_\ell^*.$$

This is the *ℓ-adic cyclotomic character of K*. This important character measures what happens to K has we adjoin more roots of unity. Extensions of K obtained by adjoining roots of unity are the *cyclotomic extensions*, and understanding them is a crucial step in understanding abelian extentions. Important point: the ℓ-adic cyclotomic character of K is a 1-dimensional representation, whereas the ℓ-adic representation of E/K is a 2-dimensional representation.

We now return to the case of elliptic curves. Observe that an isogeny $\phi : E_1 \to E_2$ induces a group homomorphism

$$\phi : E_1[\ell^n] \to E_2[\ell^n],$$

as isogenies commute with multiplication by ℓ^n map. Therefore, ϕ induces a well-defined $\mathbb{Z}_\ell$-linear map on the Tate modules,

$$\phi_\ell : T_\ell(E_1) \to T_\ell(E_2).$$

More abstractly, this says there is a homomorphism

$$\mathrm{Hom}_{\mathbb{Z}}(E_1, E_2) \to \mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(E_1), T_\ell(E_2)),$$

where $\mathrm{Hom}_{\mathbb{Z}}(E_1, E_2)$ denotes $\mathbb{Z}$-module of isogenies $E_1 \to E_2$, and $\mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(E_1), T_\ell(E_2))$ denotes the $\mathbb{Z}_\ell$-module of $\mathbb{Z}_\ell$-linear maps from $T_\ell(E_1) \to T_\ell(E_2)$. A natural question: *Why map someting that seems easy to understand, like isogenies, into something that seems more complicated to understand?* The answer is that, in fact, $\mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(E_1), T_\ell(E_2))$ can be understood using linear algebra, whereas isogenies in $\mathrm{Hom}_{\mathbb{Z}}(E_1, E_2)$ are given by horrendous formulas. Concretely, $\mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(E_1), T_\ell(E_2))$ is just given by $M_2(\mathbb{Z}_\ell)$, which means it's just a rank four $\mathbb{Z}_\ell$-module. It is fairly easy to show that this map is injective, and in fact, this map is even injective once we extend the scalars:

Theorem 4.2.3. *The map*

$$\mathrm{Hom}_{\mathbb{Z}}(E_1, E_2) \otimes_{\mathbb{Z}} \mathbb{Z}_\ell \to \mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(E_1), T_\ell(E_2)) : \phi \mapsto \phi_\ell$$

is injective.

The proof is in the book.

Corollary 4.2.4. *We have* $\mathrm{rank}(\mathrm{Hom}_{\mathbb{Z}}(E_1, E_2)) \leq 4$*, and in particular,* $\mathrm{rank}(\mathrm{End}(E)) \leq 4$*.*

Proof. If $r = \mathrm{rank}(\mathrm{Hom}_{\mathbb{Z}}(E_1, E_2))$, then because $\mathrm{Hom}_{\mathbb{Z}}(E_1, E_2)$ is a torsion free $\mathbb{Z}$-module, we have that

$$\mathrm{Hom}_{\mathbb{Z}}(E_1, E_2) \otimes_{\mathbb{Z}} \mathbb{Z}_\ell = \mathbb{Z}^r \otimes_{\mathbb{Z}} \mathbb{Z}_\ell = \mathbb{Z}_\ell^r$$

injects into

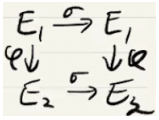
$$\mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(E_1), T_\ell(E_2)) = \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_\ell^2, \mathbb{Z}_\ell^2) = M_2(\mathbb{Z}_\ell) = \mathbb{Z}_\ell^4,$$

which has rank at most four as a $\mathbb{Z}_\ell$ module. □

Let E_1 and E_2 be elliptic curves defined over K. Then potentially E_1 and E_2 have isogenies defined over fields bigger than K. For example, $y^2 = x^3 + x$ is defined over $\mathbb{Q}$, but has an isogeny given by $(x, y) \mapsto (-x, iy)$, which is defined over $\mathbb{Q}(i)$. So it is natural to define

$$\mathrm{Hom}_K(E_1, E_2) := \{\phi \in \mathrm{Hom}(E_1, E_2) : \phi \text{ is defined over } K\}.$$

Another way to state this condition is that $\phi^\sigma = \phi$ for all $\sigma \in G_K$. On points, this means that $(\phi(P))^\sigma = \phi^\sigma(P^\sigma) = \phi(P^\sigma)$. Alternatively, ϕ is defined over K if the following diagram commutes for every element of the Galois group:



This is all to say that we could have equivalently defined

$$\mathrm{Hom}_K(E_1, E_2) := \{\phi \in \mathrm{Hom}(E_1, E_2) : \phi \circ \sigma = \sigma \circ \phi\}.$$

Therefore, if we want to do a similar construction for the Tate module, then it's natural to take as our definition

$$\begin{aligned}\mathrm{Hom}_K(T_\ell(E_1), T_\ell(E_2)) &= \{\Phi \in \mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(E_1), T_\ell(E_2)) : \Phi^\sigma = \Phi,\ \forall \sigma \in G_K\} \\ &= \mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(E_1), T_\ell(E_2))^{G_K},\end{aligned}$$

where G_K acts on $\mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(E_1), T_\ell(E_2))$ by conjugation, i.e $\Phi^\sigma := \sigma^{-1}\Phi\sigma$. And in fact, $\mathrm{Hom}_K(E_1, E_2) = \mathrm{Hom}_{\mathbb{Z}}(E_1, E_2)^{G_K}$ as well. With all that buildup:

Theorem 4.2.5. *The natural map*

$$\mathrm{Hom}_K(E_1, E_2) \otimes_{\mathbb{Z}} \mathbb{Z}_\ell \hookrightarrow \mathrm{Hom}_K(T_\ell(E_1), T_\ell(E_2))$$

is an isomorphism if:

1. *(Tate) K is a finite field;*
2. *(Faltings) K is a number field.*

The Tate module $T_\ell(E)$ contains group theoretic information about E (every ℓ^n torsion) as well number theoretic information about (with the G_K action). At the same time, $T_\ell(E)$ a linear algebraic object, as it's isomorphic to $\mathbb{Z}_\ell^2$. Going from the curve to the Tate module is directly analogous to going from a manifold to its homology; recall that this is useful because it replaces a geometric object with an algebraic one by throwing away lots of information, but not too much interesting information. Continuing the analogy further, we find it useful to look at the maps between manifolds by studying the induced maps between their homologies. So in some sense, *the Tate module is the "homology of E."* We will make this analogy precise now.

We'll see later that there is a group isomorphism

$$E(\mathbb{C}) \cong \mathbb{C}/\text{lattice} \cong \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}.$$

But the first homology group of the product of two circle groups is just

$$H_1(\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}, \mathbb{Z}) \cong \mathbb{Z} \times \mathbb{Z}$$

because of the two independent loops in the torus. Now, suppose instead that we take $\mathbb{Z}/\ell_n\mathbb{Z}$ coefficients; in this case, we know that

$$H_1(\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}, \mathbb{Z}/\ell_n\mathbb{Z}) \cong \mathbb{Z}/\ell_n\mathbb{Z} \times \mathbb{Z}/\ell_n\mathbb{Z} \cong \left(\frac{1}{\ell^n}\mathbb{Z} \times \frac{1}{\ell^n}\mathbb{Z}\right)/(\mathbb{Z} \times \mathbb{Z}),$$

which is exactly the ℓ^n torsion points in the group $\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$, which we recall is isomorphic to E. Therefore, taking the inverse limit of this gives the Tate module. So at least over $\mathbb{C}$, we have the identity

$$T_\ell(E) \cong H_1(E(\mathbb{C}), \mathbb{Z}_\ell).$$

Intuitively, we should think of the Tate module as a homology group with an action of Galois, from which we'll be able to deduce tons of information about the elliptic curve.

In the next lecture we'll look at an alternating bilinear pairing on the Tate module, which will essentially be the wedge product, but it'll be defined intrinsically so it'll interact well with the actiion of the Galois.

(Lecture 17: March 1, 2021)

4.3 The Weil pairing

Recall that

$$E[m] := \{P \in E(\overline{K}) : [m]P = \mathcal{O}\}$$

is isomorphic to $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$ when $\operatorname{char} K = 0$ or $\operatorname{char} K \nmid m$. We can think of this as a 2-dimensional vector space, so there ought to be an alternating bilinear nondegenerate pairing that maps to the scalars,

$$E[m] \times E[m] \to \mathbb{Z}/m\mathbb{Z}.$$

Today we will define such a pairing intrinsically, and in a way that respects the Galois action. Throughout the construction, we'll use the following important fact:

Fact 4.3.1. Given a divisor $D = \sum n_P(P)$ on E, $D \sim 0$ if and only if the following are true:

1. $\sum n_P = 0$,
2. $\sum [n_P]P = \mathcal{O}$.

Suppose $T \in E[m]$, and consider the divisor $m(T) - m(\mathcal{O})$. This divisor has degree 0 and adds to $\mathcal{O}$, so there exists some rational function $f_T \in \overline{K}(E)$ such that

$$\operatorname{div}(f_T) = m(T) - m(\mathcal{O}).$$

Note that such an f_T is unique up to multiplication by a nonzero scalar. Next, choose any $T' \in E$ with $[m]T' = T$; such a T' exists because $[m] : E \to E$ is surjective. Then we can compute that

$$D := [m]^*(T) - [m]^*(\mathcal{O}) = \sum_{P \in [m]^{-1}(T)} (P) - \sum_{P \in [m]^{-1}(\mathcal{O})} (P) = \sum_{S \in E[m]} (T' + S) + \sum_{S \in E[m]} (S),$$

because $[m]$ is separable and unramified.

Claim 4.3.2. $D \sim 0$, and in particular, there exists $g_T \in \overline{K}(E)$ with

$$\operatorname{div}(g_T) = [m]^*(T) - [m]^*(\mathcal{O}).$$

Proof. We know

$$\deg D = \deg\left(\sum_{S\in E[m]}(T'+S) - \sum_{S\in E[m]}(S)\right) = \deg\left(\sum_{S\in E[m]}(T'+S)\right) - \deg\left(\sum_{S\in E[m]}(S)\right) = 0,$$

as each sum has $\#E[m]$ terms. Summing this divisor in E, we get

$$\operatorname{sum}_E(D) = \sum_{S\in E[m]}\{T'+S\} - \sum_{S\in E[m]} S = (\#E[m])\cdot T' = [m^2]T' = [m]T = \mathcal{O},$$

as needed. □

Next, we know that the divisor of a composition can be computed as follows:

$$\operatorname{div}(f_T \circ [m]) = [m]^*\operatorname{div}(f_T) = [m]^*(m(T) - m(\mathcal{O})) = m([m]^*(T) - [m]^*(\mathcal{O})) = m\cdot\operatorname{div}(g_T) = \operatorname{div}(g_T^m).$$

This means that $f_T \circ [m]/g_T^m$ has no poles or zeroes, and is therefore constant, hence $f_T \circ [m] = c\cdot g_T^m$ for some $c \in \overline{K}^*$. If we replace f_T with $c^{-1}f_T$, then we get

$$f_T \circ [m] = g_T^m.$$

Recapitulating this construction: We started with $T \in E[m]$ and obtained a rational function $f_T \in \overline{K}(E)$ with $\operatorname{div}(f_T) = m(T) - m(\mathcal{O})$. We chose some T' with $[m](T') = T$, and used this to obtain a rational function $g_T \in \overline{K}(E)$ with $\operatorname{div}(g_T) = [m]^*(T) - [m]^*(\mathcal{O})$. We then normalized these functions in such a way that $f_T \circ [m] = g_T^m$. These functions are the basic building blocks for the Weil pairing, which we can now describe.

Let $S \in E[m]$ be any m-torsion point, and consider the rational map

$$G_{S,T} : E \to \mathbb{P}^1 : P \mapsto \frac{g_T(P+S)}{g_T(P)}.$$

We can compute that

$$G_{S,T}(P)^m = \frac{g_T(P+S)^m}{g_T(P)^m} = \frac{f_T([m](P+S))}{f_T([m](P))} = \frac{f_T([m](P))}{f_T([m](P))} = 1,$$

for generic $P \in E$. In other words, $G_{S,T} : E \to \mathbb{P}^1$ is a rational function with the property that

$$G_{S,T}(P) \in \mu_m, \qquad \text{for generic } P \in E.$$

In particular, $G_{S,T}$ is not surjective. As every morphism between curves is either surjective or constant, it follows that $G_{S,T}$ is constant, and its only value is in some m'th root of unity. In fact, this value is the quantity we're interested in.

Definition 4.3.3. The *Weil pairing* on $E[m]$ is the map

$$e_m : E[m]\times E[m] \to \mu_m : e_m(S,T) = \frac{g_T(P+S)}{g_T(P)}, \qquad \text{for any } P \in E.$$

Concretely, here is the algorithm for computing the Weil pairing:

- Begin with $(S,T) \in E[m]\times E[m]$.
- Obtain rational functions $f_T, g_T \in \overline{K}(E)$ satisfying $\operatorname{div}(f_T) = m(T) - m(\mathcal{O})$ and $\operatorname{div}(g_T) = [m]^*(T) - [m]^*(\mathcal{O})$ such that $f_T \circ [m] = g_T^m$.

- We define $e_m(S,T) = \frac{g_T(P+S)}{g_T(P)}$ for any $P \in E$ where both $g_T(P+S), g_T(P) \neq 0, \infty$.

We proved that the value $e_m(S,T)$ is an m-th root of unity that depends on S and T. And this is well-defined, as the only choice that we made is the constant multiple of g_T that we chose. But if we replace g_T with $c \cdot g_T$, then we can see the constant cancels in the quotient.

Theorem 4.3.4. *The Weil pairing satisfies the following properties:*

(a) e_m *is bilinear.*

(b) e_m *is alternating, meaning* $e_m(T,T) = 1$. *So by bilinearity,* $e_m(S,T) = e_m(T,S)^{-1}$.

(c) e_m *is Galois invariant, meaning for* $\sigma \in G(\overline{K}/K)$, *we have* $e_m(S,T)^\sigma = e_m(S^\sigma, T^\sigma)$.

Proof. For linearity in the first coordinate, let $R, S, T \in E[m]$, and P a generic point on E. We compute

$$e_m(R+S,T) = \frac{g_T(P+R+S)}{g_T(P)} = \frac{g_T(P+R+S)}{g_T(P+R)}\frac{g_T(P+R)}{g_T(P)} = e_m(S,T)e_m(R,T).$$

For linearity in the second coordinate, take $T_1, T_2 \in E[m]$ and $T_3 = T_1 + T_2$. Let $f_i = f_{T_1}$ and $g_i = g_{T_i}$. The key is to use the addition laws with divisors to look at the divisor

$$(T_1 + T_2) - (T_1) - (T_2) + (\mathcal{O}).$$

This clearly has degree 0 and sums to $\mathcal{O}$ in E, so it is the divisor of some rational function (h), so we can compute

$$\begin{aligned}\operatorname{div}\left(\frac{f_3}{f_1 f_2}\right) &= m(T_1+T_2) - m(\mathcal{O}) - m(T_1) + m(\mathcal{O}) - m(T_2) + m(\mathcal{O}) \\ &= m[(T_1+T_2) - (T_1) - (T_2) + (\mathcal{O})] \\ &= m \cdot \operatorname{div}(h),\end{aligned}$$

which implies that $f_3 = c f_1 f_2 h^m$. On the other hand, we have

$$g_3^m = f_3 \circ [m] = (c f_1 f_2 h^m) \circ [m] = c(f_1 \circ [m])(f_2 \circ [m])(h \circ [m])^m = c g_1^m g_2^m (h \circ [m])^m,$$

thus $g_3 = c' g_1 g_2 (h \circ m)$ for some $c' \in \overline{K}^*$. So we can compute the Weil pairing

$$e_m(S, T_1+T_2) = \frac{g_3(P+S)}{g_3(P)} = \frac{c' g_1(P+S) g_2(P+S) h([m]P + [m]S)}{c' g_1(P) g_2(P) h([m]P)} = e_m(S,T_1)e_m(S,T_2) \cdot 1$$

because $[m]S = \mathcal{O}$. This proves (a) by induction. And (c) is clear because all the functions we're using were rational functions with coefficients in $\overline{K}$.

(Lecture 18: March 3, 2021)

Now let's show that e_m is alternating. Recall the translation map $\tau_P : E \to E : Q \mapsto Q + P$. Because $\operatorname{div}(f_T) = m(T) - m(\mathcal{O})$, we can compute

$$\operatorname{div}\left(\prod_{j=0}^{m-1} f_T \circ \tau_{[j]T}\right) = \sum_{j=0}^{m-1} \operatorname{div}(f_T \circ \tau_{[j]T}) = \sum_{j=0}^{m-1} m([1-j]T) - m([-j]T),$$

because the composition $f_T \circ \tau_{[j]T}$ just translates the divisors of f_T by $[j]T$. But this sum telecopes, and in fact it vanishes completely, as $[m]T = \mathcal{O}$. It follows that

$$\prod_{j=0}^{m-1} f_T \circ \tau_{[j]T} = \text{constant}.$$

This implies that

$$\prod_{j=0}^{m-1} g_T \circ \tau_{[j]T'} = \text{constant}$$

because

$$\left(\prod_{j=0}^{m-1} g_T \circ \tau_{[j]T'}\right)^m = \prod_{j=0}^{m-1} f_T \circ [m] \circ \tau_{[j]T'} = \prod_{j=0}^{m-1} f_T \circ \tau_{[j][m]T'} \circ [m] = \left(\prod_{j=0}^{m-1} f_T \circ \tau_{[j]T}\right) \circ [m].$$

What we're going to do is evaluate this constant function at P and $P + T'$. On the one hand, we have

$$\prod_{j=0}^{m-1} g_T \circ \tau_{[j]T'}(P) = \prod_{j=0}^{m-1} g_T(P + [j]T'),$$

and on the other hand

$$\prod_{j=0}^{m-1} g_T \circ \tau_{[j]T'}(P + T') = \prod_{j=0}^{m-1} g_T(P + [j+1]T').$$

If we divide out the common terms, then we get that the $j = 0$ term on the first equation equals the $j = m-1$ term on the second equation, which implies that

$$g_T(P) = g_T(P + [m]T') = g_T(P + T),$$

so $e_m(T, T) = g_T(P + T)/g_T(P) = 1$, as needed. □

Remark 4.3.5. Why do we need f_T at all in the definition of the Weil pairing? We could have easily defined the Weil pairing without defining f_T at all, by just taking g_T so that $(g_T) = [m]^*(T) - [m]^*(\mathcal{O})$. But whenever we're proving something about the Weil pairing, we use crucially that $g_T^m = f_T \circ [m]$, such as when we showed e_m is alternating.

Because an isogeny between elliptic curves maps $E_1[m]$ to $E_2[m]$, a result of the following form is natural to ask for:

Theorem 4.3.6. *Let $\phi : E_1 \to E_2$ be an isogeny. Given $S \in E_1[m]$ and $T \in E_2[m]$, we have*

$$e_{m,E_2}(\phi(S), T) = e_{m,E_1}(S, \hat{\phi}(T)).$$

Why is this formula reasonable? If m is prime, then $\phi : E_1[m] \to E_2[m]$ is an $\mathbb{F}_p$-linear transformation of $\mathbb{F}_p$ vector spaces, so it has some adjoint with respect to the linear pairing e_m. This theorem says that the dual isogeny $\hat{\phi}$ is in fact the adjoint for the Weil pairing.

Proof. It suffices to show that the dual isogeny is the adjoint for the Weil pairing. By definition,

$$e_m(\phi(S), T) = \frac{g_T(P + \phi(S))}{g_T(P)}.$$

On the other hand, we can compute

$$\phi^*((T) - (\mathcal{O})) = \sum_{P \in \phi^{-1}(T)} (P) - \sum_{P \in \phi^{-1}(\mathcal{O})} (P).$$

Observe that the points in this divisor actually add to $\hat{\phi}(T)$; this is because $\hat{\phi}$ is the unique isogeny which satisfies $\phi \circ \hat{\phi} = [\deg \phi]$, and

$$\sum_{P \in \phi^{-1}(T)} \phi(P) - \sum_{P \in \phi^{-1}(\mathcal{O})} \phi(P) = [\deg \phi]T.$$

This implies that that the divisors $\phi^*[(T)-(\mathcal{O})]$ and $(\hat{\phi}(T))-(\mathcal{O})$ sum to the same point, $\hat{\phi}(T)$; that is, the difference of these divisors has degree zero and the constituent points sum to $\mathcal{O}$, so there exists a rational function h so that

$$\phi^*[(T)-(\mathcal{O})] = (\hat{\phi}(T))-(\mathcal{O}) + \mathrm{div}(h).$$

We compute that

$$\begin{aligned}\mathrm{div}\left(\frac{f_T\circ\phi}{h^m}\right) &= \phi^*\mathrm{div}(f_T) - m\cdot\mathrm{div}(h)\\ &= \phi^*(m(T)-m(\mathcal{O})) - m(\phi^*[(T)-(\mathcal{O})]-(\hat{\phi}(T))+(\mathcal{O}))\\ &= m\hat{\phi}(T) - m(\mathcal{O}),\end{aligned}$$

so $f_{\hat{\phi}(T)} = \frac{f_T\circ\phi}{h^m}$. On the other hand, we compute

$$\left(\frac{g_T\circ\phi}{h\circ[m]}\right)^m = \frac{f_T\circ[m]\circ\phi}{(h\circ[m])^m} = \left(\frac{f_T\circ\phi}{h^m}\right)\circ[m],$$

so $g_{\hat{\phi}(T)} = \frac{g_T\circ\phi}{h\circ[m]}$ satisfies $f_{\hat{\phi}(T)}\circ[m] = g^m_{\hat{\phi}(T)}$. Therefore, the definition of the Weil pairing implies that

$$e_m(s,\hat{\phi}(T)) = \frac{g_{\hat{\phi}(T)}(P+S)}{g_{\hat{\phi}(T)}(P)} = \frac{\frac{g_T\circ\phi}{h\circ[m]}(P+S)}{\frac{g_T\circ\phi}{h\circ[m]}(P)} = \frac{g_T(\phi(P)+\phi(S))}{g_T(\phi(P))}\cdot\frac{h(P)}{h(P)} = e_m(\phi(S),T).$$

This completes the proof. □

Next, we would like to fit together the maps

$$e_m : E[m]\times E[m]\to\mu_m.$$

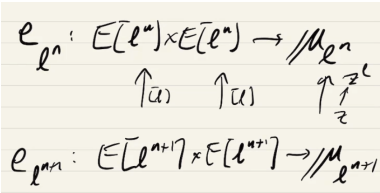
to obtain something in characteristic zero. So what we do is take the pairing on the ℓ^n torsion,

$$e_{\ell^n} : E[\ell^n]\times E[\ell^n]\to\mu_{\ell^n}$$

and the pairing on the ℓ^{n+1} torsion

$$e_{\ell^{n+1}} : E[\ell^{n+1}]\times E[\ell^{n+1}]\to\mu_{\ell^{n+1}}$$

and connect these pairings using the multiplication by ℓ map:



Once can show that diagram commutes, i.e., that there is a compatibility between the Weil pairing at the different levels of torsion. So we can take the inverse limit and obtain a pairing on the inverse limits:

Theorem 4.3.7. *There exists a bilinear, alternating, nondegenerate, Galois-invariant pairing*

$$e_\ell : T_\ell(E)\times T_\ell(E)\to T_\ell(\mu).$$

Further, given an isogeny $\phi : E_1\to E_2$, its adjoint with respect to this pairing is the dual, i.e.

$$e_\ell(\phi x, y) = e(x,\hat{\phi}y).$$

Sometimes we write $e = e_\ell$. If you choose bases for all the spaces in $T_\ell(E) \times T_\ell(E) \to T_\ell(\mu)$, then we get a map

$$e_\ell : \mathbb{Z}_\ell^2 \times \mathbb{Z}_\ell^2 \to \mathbb{Z}_\ell.$$

Crucially, the target space $\mathbb{Z}_\ell$ comes equipped with the Galois action on the roots of unity. And if we want to be completely precise about this pairing: e_ℓ gives a Galois invariant isomorphism between the second alternating tensor product to the Tate module of the roots of unity,

$$e_\ell : \bigwedge{}^2 T_\ell(E) \to T_\ell(\mu).$$

The fact that e_ℓ respects Galois is super important; it lets us respect arithmetic.

Proposition 4.3.8. *Let $\phi : E \to E$ be an isogeny, so ϕ induces a map $\phi_\ell : T_\ell(E) \to T_\ell(E)$. As $T_\ell(E)$ is a rank two $\mathbb{Z}_\ell$-module, ϕ_ℓ has a determinant map* $\det : T_\ell(E) \to \mathbb{Z}_\ell$ *and trace map* $\operatorname{tr} : T_\ell(E) \to \mathbb{Z}_\ell$. *Then:*

(a) $\det(\phi_\ell) = \deg(\phi)$,

(b) $\operatorname{tr}(\phi_\ell) = 1 + \deg\phi - \deg(1 - \phi)$.

In particular, the determinant and trace maps are independent of ℓ.

Proof. For (a), choose a basis v_1, v_2 for $T_\ell(E)$. Then $\phi_\ell v_1 = av_1 + bv_2$ and $\phi_\ell v_2 = cv_1 + dv_2$, where $a, b, c, d \in \mathbb{Z}_\ell$. Then we can compute

$$e(v_1, v_2)^{\deg\phi} = e((\deg\phi) \cdot v_1, v_2) = e(\hat{\phi}_\ell \phi_\ell v_1, v_2) = e(\phi_\ell v_1, \phi_\ell v_2) = e(av_1 + bv_2, cv_1 + dv_2) = e(v_1, v_2)^{ad-bc}.$$

As e is nondegenerate, we conclude that

$$\deg\phi = ad - bc = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det \phi_\ell.$$

Finally, note that (b) follows from (a) and just working with 2×2 matrices. □

(Lecture 19: March 5, 2021)

Recall that the Frobenius map $\phi_q : E \to E$ satisfies

$$\deg(1 - \phi_q) = \#\ker(1 - \phi_q) = \#\{P \in E : \phi_q(P) = P\} = \#E(\mathbb{F}_q).$$

On the other hand, we can compute the degree by computing the determinant of the action of the Tate module. We'll use this later when we prove the Weil conjectures for elliptic curves. We'll first consider the general case.

4.4 The Weil conjectures

Consider any nonsingular projective variety $V \subseteq \mathbb{P}^n$ defined over $\mathbb{F}_q$. Concretely, V is cut out by $f_1, \dots, f_r \in \mathbb{F}_q[x_1, \dots, x_n]$. Weil asked:

How does $\#V(\mathbb{F}_{q^n})$ behave as n varies?

It is natural to combine these point counts into a generating function; we do this with arithmetically interesting sequences all the time.

Example 4.4.1. If F_n is the n'th Fibonacci number, then

$$\sum_{n=0}^{\infty} F_n x^n = \frac{1}{1 - x - x^2}.$$

This can be proved by expanding the RHS using partial fractions and expanding the geometric series that result.

Definition 4.4.2. Let (a_n) be some real sequence. We associate three generating functions to (a_n).

1. The *geometric* generating function is $\sum_{n\geq 0} a_n x^n$, so called because $a_n \equiv 1$ gives $1/(1-x)$.
2. The *exponential* generating function is $\sum_{n\geq 0} \frac{a_n}{n!} x^n$, so called because $a_n \equiv 1$ gives $\exp(x)$.
3. The *logarithmic* generating function is $\sum_{n\geq 1} \frac{a_n}{n} x^n$, so called because $a_n \equiv 1$ gives $-\log(1-x)$.

When analyzing a sequence, it is natural to construct each of these natural generating functions until you find one that has nice enough properties to work with.

Definition 4.4.3. The *zeta function* of $V/\mathbb{F}_q$ is

$$Z(V/\mathbb{F}_q, T) := \exp\left(\sum_{n\geq 1} \frac{\#V(\mathbb{F}_{q^n})}{n} T^n\right) \in \mathbb{Q}[[T]].$$

Often, when you prove things about generating functions, you can recover interesting information about the sequence you fed into it; for example,

$$\frac{1}{(n-1)!}\frac{d^n}{dT^n} \log Z(V/F_q, T)\Big|_{T=0} = \#\mathbb{F}_q(V).$$

Example 4.4.4. Take $V = \mathbb{P}^N$. Then we can compute

$$\#\mathbb{P}^N(\mathbb{F}_{q^n}) = \#\left(\frac{(\mathbb{F}_{q^n})^{N+1} - \{0\}}{\mathbb{F}_{q^n}^*}\right) = \frac{(q^n)^{N+1}-1}{q^n - 1} = \sum_{i=0}^{N} (q^n)^i.$$

We have a good handle on these numbers, so we don't need to put them into a zeta function to understand them; but to illustrate the zeta function we'll do this anyway. We compute

$$\log Z(\mathbb{P}^N/\mathbb{F}_q, T) = \sum_{n\geq 1}\sum_{i=0}^{N} (q^n)^i \frac{T^n}{n} = \sum_{i=0}^{N}\sum_{n=1}^{\infty} \frac{(q^i T)^n}{n} = \sum_{i=0}^{N} -\log(1-q^iT) = \log \prod_{i=0}^{N} (1-q^iT)^{-1}.$$

Note that because we're working in the power series ring $\mathbb{Q}[[T]]$, we can rearrange infinite series at will, since we don't care about convergence. In summary, we've shown that

$$Z(\mathbb{P}^n/\mathbb{F}_q, T) = \prod_{i=0}^{n} \frac{1}{1-q^iT}.$$

A priori, $Z(\mathbb{P}^n/\mathbb{F}_q, T)$ is a power series; but we've shown that in fact $Z(\mathbb{P}^n/\mathbb{F}_q, T)$ is a rational function. Knowing that a power series is a rational function tells us that the coefficients behave in a very well-defined manner; for example, such a power series has coefficients that satisfy a recurrence.

Theorem 4.4.5 (Weil conjectures)**.** *Let $V/\mathbb{F}_q$ be a smooth projective variety of dimension N.*

(a) (Rationality:) *$Z(V/\mathbb{F}_q, T) \in \mathbb{Q}(T)$.*

(b) (Functional equation:) *There exists $\chi \in \mathbb{N}$ (called the* Euler characteristic *of V) such that*

$$Z\left(V/\mathbb{F}_q, \frac{1}{q^N T}\right) = \pm q^{\frac{N\chi}{2}} T^\chi Z(V/\mathbb{F}_q, T).$$

(c) (Riemann hypothesis:) *we have*

$$Z(V/\mathbb{F}_q, T) = \frac{P_1(T)P_3(T)\cdots P_{2N-1}(T)}{P_0(T)P_2(T)\cdots P_{2N(T)}},$$

where each $P_i(T) \in \mathbb{Z}[T]$, and each of these polynomials factors as

$$P_i(T) = \prod_{j=1}^{b_i} (1-\alpha_{ij}T)$$

with $|\alpha_{ij}| = q^{i/2}$. Furthermore, $P_0 = 1-T$ and $P_{2N}(T) = 1-q^nT$.

Weil made these conjectures in 1949. He proved them in the case where V is a curve, as well as in the case of abelian varieties. The rationality of the zeta function was proven by Dwork in 1960, using p-adic functional analysis. Then in 1960, Grothendieck started a decade long program to prove the Weil conjectures. This program required re-tooling algebraic geometry, developing schemes and sheaves in the process. They also developed a cohomology theory. In 1965, Grothendieck proved the functional equation, and in 1974, Deligne used the cohomology theory to prove the Riemann hypothesis, which was the hardest part of the conjectures.

Now we will turn to using the Weil conjectures to estimate point coints. Consider the zeta function

$$Z(V/\mathbb{F}_q, T) = \exp\left(\sum_{n\geq 1} \frac{\#V(F_{q^n})}{n} T^n\right).$$

We take the logarithm of this and apply the Weil conjectures, yielding

$$\sum_{n=1}^{\infty} \frac{\#V(\mathbb{F}_{q^n})}{n} T^n = \sum_{i=0}^{2N} (-1)^{i+1} \sum_{j=1}^{b_i} \log(1-\alpha_{ij}T) = \sum_{i=0}^{2N} (-1)^i \sum_{j=1}^{b_i} \sum_{n=1}^{\infty} \frac{(\alpha_{ij}T)^n}{n} = \sum_{n=1}^{\infty} \sum_{i=0}^{2N} (-1)^i \sum_{j=1}^{b_i} \alpha_{ij}^n \frac{T^n}{n}.$$

If we compare the coefficients of T^n, then we obtain the identity

$$\#V(\mathbb{F}_{q^n}) = \sum_{i=0}^{2N} (-1)^i \sum_{j=1}^{b_i} \alpha_{ij}^n, \qquad \text{where } |\alpha_{ij}| = q^{i/2}.$$

This is quite simple as a function of n; it's a finite sum of some fixed quantities, raised to the n'th power. Observe that when $i = 2n$ the α_{ij} are biggest, and when when $i = 0$ the α_{ij} are smallest. Using this, as well as the fact that the last polynomial is $P_{2N}(T) = 1 - q^N T$, one can show that

$$\#V(\mathbb{F}_{q^n}) = q^{Nn} - \sum b_i \text{ terms, each with absolute values } \leq q^{\frac{n(2N-1)}{2}} = q^{n(N-\frac{1}{2})} + \text{other terms.}$$

This implies in particular that

$$|\#V(\mathbb{F}_{q^n}) - q^{Nn}| \leq C_{V/\mathbb{F}_q}.$$

Note that the b_i's depend only on $V/\mathbb{F}_q$; in fact, the b_i are the Betti numbers

$$b_i = \dim H^i_{et}((V/\mathbb{F}_q), \mathbb{Q}_\ell).$$

(Lecture 20: March 8, 2021)

4.5 Proof of the Weil conjectures for elliptic curves

Recall that an elliptic curve endomorphism $\psi \in \mathrm{End}(E)$ induces a Tate module endomorphism $\psi_\ell \in \mathrm{End}(T_\ell(E))$ which satisfies $\det \psi_\ell = \deg \phi$ and $\operatorname{tr} \psi_\ell = 1 + \deg \psi - \deg(1-\psi)$.

Let $E/\mathbb{F}_q$ be an elliptic cirve, and let $\phi = \phi_q = (x^q, y^q)$ be the Frobenius map. Then $\phi^n = \phi_{q^n}$. *How many points does E have over $\mathbb{F}_{q^n}$?* This quantity is just

$$\#E(\mathbb{F}_{q^n}) = \deg(1-\phi_q^n) = 1 + \deg(\phi_q^n) - \operatorname{tr}(\phi_q^n \curvearrowright T_\ell(E)).$$

What is $\operatorname{tr}(\phi_q^n \curvearrowright T_\ell(E))$*?* If a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has characteristic polynomial $T^2 - (\operatorname{tr} A)T + \det A = (T-\alpha)(T-\beta)$, so A has eigenvalues α and β, then it's a general fact that $\operatorname{tr}(A^n) = \alpha^n + \beta^n$. This implies that

$$\#E(\mathbb{F}_{q^n}) = q^n - \alpha_q^n - \beta_q^n + 1, \tag{4.1}$$

where α_q and β_q satsify $\#E(\mathbb{F}_q) = q - \alpha_q - \beta_q + 1$, and where α_q and β_q are the roots of

$$T^2 - \operatorname{tr}(\phi_q \curvearrowright T_\ell(E))T + \det(\phi_q \curvearrowright T_\ell(E)) = T^2 - (\deg \phi_q + 1 - \deg(1-\phi_q))T + q = (T-\alpha_q)(T-\beta_q).$$

On the other hand, we know that

$$\deg \phi_q + 1 - \deg(1 - \phi_q) = q + 1 - \#E(\mathbb{F}_q).$$

Thus, given an elliptic curve E, the characteristic polynomial of $\phi_{q^n} \curvearrowright T_\ell(E)$ can be explicitly computed as soon as we know how many points E has over $\mathbb{F}_q$, as the latter computation directly yields α_q and β_q.

Definition 4.5.1. The quantity $\deg \phi_q + 1 - \deg(1 - \phi_q) = q + 1 - \#E(\mathbb{F}_q)$ is called the *trace of Frobenius*, as it is equal to $\operatorname{tr}(\phi_q \curvearrowright T_\ell(E)) = \alpha_q + \beta_q$. It's commonly denoted a_q.

As $\alpha_q \beta_q = q$, and we know that the quadratic characteristic polynomial has complex roots, the roots must be complex conjugates, so

$$|\alpha_q| = |\beta_q| = \sqrt{q}.$$

We are now ready to prove the Weil conjectures for elliptic curves. Using (4.1), we can compute that the zeta function is

$$\begin{aligned}
\log Z(E/\mathbb{F}_q, T) &= \sum_{n \geq 1} \frac{\#E(F_{q^n})}{n} T^n \\
&= \sum_{n \geq 1} \frac{(qT)^n}{n} - \sum_{n \geq 1} \frac{(\alpha_q T)^n}{n} - \sum_{n \geq 1} \frac{(\beta_q T)^n}{n} + \sum_{n \geq 1} \frac{T^n}{n} \\
&= -\log(1 - qT) + \log(1 - \alpha_q T) + \log(1 - \beta_q T) - \log(1 - T),
\end{aligned}$$

thus

$$Z(E/\mathbb{F}_q, T) = \frac{(1 - \alpha_q T)(1 - \beta_q T)}{(1 - T)(1 - qT)} = \frac{1 - a_q T + qT^2}{(1 - T)(1 - qT)},$$

where $a_q = q + 1 - \#E(\mathbb{F}_q)$ is the trace of Frobenius.

Recall that the estimate $|\alpha| = |\beta| = q^{1/2}$ is called the "Riemann hypothesis for elliptic curves over finite fields." Upon first glance, this doesn't look much like the classical Riemann hypothesis, which says the roots of the Riemann zeta function are on the line $\Re(s) = 1/2$. How do we see the analogy? Instead of using T as the variable, we use q^{-s} as the variable, where $s \in \mathbb{C}$. Then we define

$$\zeta_{E/\mathbb{F}_q}(s) := Z(E/\mathbb{F}_q, q^{-s}) = \frac{1 - a_q q^{-s} - q^{1-2s}}{(1 - q^{-s})(1 - q^{1-s})}.$$

We know

$$\zeta_{E/\mathbb{F}_q}(s) = 0 \iff q^{-s} \text{ is a root of } 1 - a_q T + qT^2 \iff q^{-s} \in \{\alpha_q, \beta_q\} \implies |q^{-s}| = q^{1/2},$$

which implies that $\Re(s) = -1/2$. So up to a change of variables, it's exactly the classical Riemann hypothesis.

Let us consider an elliptic curve $E/\mathbb{Q}$. It is natural to look at

$$\zeta(E/\mathbb{Q}, s) := \prod_p \zeta(\tilde{E}_p/\mathbb{F}_p, s) = \prod_p \frac{1 - a_p p^{-s} + p^{1-2s}}{(1 - p^{-s})(1 - p^{1-s})} = \zeta(s)\zeta(1 - s) \prod_p (1 - a_p p^{-s} + p^{1-2s}).$$

As the numerator is the only aspect of the Euler product that is specific to the elliptic curve E, it is natural to define[10]

$$L(E/\mathbb{Q}, s) := \prod_p \frac{1}{1 - a_p p^{-s} + p^{1-2s}}.$$

One can show that Hasse's estimate $|a_p| \leq 2\sqrt{p}$ implies this converges for $\Re(s) > 3/2$. In fact:

Theorem 4.5.2 (Wiles)**.** *$L(E/\mathbb{Q}, s)$ extends to a holomorphic function on $\mathbb{C}$.*

Note that if you replace $\mathbb{Q}$ by a number field, in general it's still an open question.

[10] We're lying a little bit; this definition changes at finitely many primes.

4.6 Exercises

Exercise (Silverman 3.16). Let E be an elliptic curve. We define a pairing

$$\tilde{e}_m : E[m] \times E[m] \to \mu_m$$

as follows: let $P, Q \in E[m]$ and choose divisors D_P and D_Q in $\mathrm{Div}^0(E)$ that add to P and Q, respectively, i.e., such that $\sigma(D_P) = P$ and $\sigma(D_Q) = Q$, where $\sigma : \mathrm{Div}^0(E) \to E$ maps a degree 0 divisor D to the unique point $P \in E$ satisfying $D \sim (P) - (\mathcal{O})$. Assume further that D_P and D_Q are chosen with disjoint supports. Since P and Q have order m, the degree zero divisors mD_P and mD_Q add to $\mathcal{O}$, so there are functions $f_P, f_Q \in \overline{K}(E)$ satisfying

$$\mathrm{div}(f_P) = mD_P \qquad \text{and} \qquad \mathrm{div}(f_Q) = mD_Q.$$

We define

$$\tilde{e}_m = \frac{f_P(D_Q)}{f_Q(D_P)}.$$

(a) Prove that $\tilde{e}_m(P, Q)$ is well-defined, i.e., its value depends only on P and Q, independent of the various choices of D_P, D_Q, f_P, and f_Q.

(b) Prove that $\tilde{e}_m(P, Q) \in \mu_m$.

(c) Prove that $\tilde{e}_m(P, Q) = e_m(Q, P)$, and hence that $\tilde{e}_m = e_m^{-1}$, where e_m is the Weil pairing.

Proof. Let f_P, f_Q, D_P, D_Q and f'_P, f'_Q, D'_P, D'_Q be two sets of choices. As D_Q and D'_Q are degree zero divisors which sum to $\mathcal{O}$, we have that $D_Q - D'_Q = \mathrm{div}(g)$ for some $g \in \overline{K}(E)$, and similarly, $D_P - D'_P = \mathrm{div}(h)$ for $h \in \overline{K}(E)$. By construction, we know

$$\mathrm{div}\left(\frac{f_P}{f'_P}\right) = m(D_P - D'_P) = m \cdot \mathrm{div}(h) = \mathrm{div}(h^m),$$

which implies that $f_P = c \cdot f'_P h^m$. The evaluation of a function at a degree zero divisor is well-defined up to the function being multiplied by a nonzero scalar, so we can compute that

$$\frac{f_P(D_Q)}{f'_P(D'_Q)} = \frac{f'_P(D_Q)h^m(D_Q)}{f'_P(D'_Q)} = f'_P(D_Q - D'_Q)h^m(D_Q) = f'_P(\mathrm{div}(g))h^m(D_Q) = g(\mathrm{div}(f'_P))h^m(D_Q).$$

Similarly, we can compute that

$$\frac{f_Q(D_P)}{f'_Q(D'_P)} = h(\mathrm{div}(f'_Q))g^m(D_P).$$

These two equations imply that

$$\frac{f_P(D_Q)/f_Q(D_P)}{f'_P(D'_Q)/f'_Q(D'_P)} = \frac{g(\mathrm{div} f'_P)}{g(D_P)^m} \cdot \frac{h(D_Q)^m}{h(\mathrm{div}(f'_Q))}.$$

But we can compute that $g(D_P)^m = g(mD_P) = g(\mathrm{div}(f_P))$, hence

$$\frac{g(\mathrm{div}(f'_P))}{g(D_P)^m} = g(\mathrm{div}(f'_P) - \mathrm{div}(f_P)) = g(m(D_P - D'_P)) = g(\mathrm{div}(h))^m,$$

and similarly $h(D_Q)^m / h(\mathrm{div}(f'_Q)) = h(\mathrm{div}(g))^{-m}$, so we're done by Weil reciprocity. This proves part (a). For part (b), simply compute that

$$\tilde{e}_m(P, Q)^m = \frac{f_P(mD_Q)}{f_Q(mD_P)} = \frac{f_P(\mathrm{div}(f_Q))}{f_Q(\mathrm{div}(f_P))} = 1$$

using Weil reciprocity. And part (c) is worked out in detail on pp. 462 of the text. □

Exercise (Silverman 5.3). Let A be a square matrix with coefficients in a field. Prove that

$$\exp\left(\sum_{n=1}^{\infty}\frac{(\operatorname{tr}A^n)T^n}{n}\right)=\frac{1}{\det(I-AT)}.$$

Proof. We will first prove this in the case where $\det A\neq 0$. The characteristic polynomial of A is $\det(\lambda I-A)=\lambda^n\det(I-A\lambda^{-1})$. If we let $T=\lambda^{-1}$, then we can see that T is a root of $\det(I-AT)$ if and only if T^{-1} is an eigenvalue of A. Say A is an $m\times m$ matrix, and $\lambda_1,\ldots,\lambda_m$ are the eigenvalues of A, counted with multiplicity. By our variable transformation, we know

$$\begin{aligned}1&=\text{leading coefficient of }\det(\lambda I-A)\in\mathbb{F}[\lambda]\\&=\text{constant coefficient of }\det(I-AT)\in\mathbb{F}[T],\end{aligned}$$

therefore we know that $\det(I-AT)=\prod_{i=1}^m(I-\lambda_iT)$, so we can compute

$$\log\left(\frac{1}{\det(1-AT)}\right)=-\sum_{i=1}^m\log(1-\lambda_iT)=\sum_{i=1}^m\sum_{n=1}^{\infty}\frac{(\lambda_iT)^n}{n}=\sum_{n=1}^{\infty}\frac{\sum_{i=1}^m\lambda_i^n}{n}T^n=\sum_{n=1}^{\infty}\frac{(\operatorname{tr}A^n)T^n}{n},$$

as needed. And the case where $\det A=0$, notice that the contribution from all the zero eigenvalues can be ignored in the above computation. □

Exercise (Silverman 5.13). Let $E/\mathbb{F}_q$ be an elliptic curve, and for each $n\geq 1$, let

$$a_n=q^n+1-\#E(\mathbb{F}_{q^n}).$$

(By convention, we set $a_0=2$.) Prove that

$$a_{n+2}=a_1a_{n+1}-qa_n\qquad\text{for all }n\geq 0.$$

This linear recurrence gives a way to compute a_n from the initial values $a_0=0$ and $a_1=q+1-\#E(F_q)$.

Proof. Recall that the trace of the n'th power of Frobenius is

$$\operatorname{tr}(\phi_q^n\curvearrowright T_\ell(E))=\alpha^n+\beta^n=q^n+1-\#E(\mathbb{F}_q^n),$$

where α and β are the eigenvalues of $\phi_q\curvearrowright T_\ell(E)$, and they satisfy $\alpha\beta=q$. So we can compute that

$$\begin{aligned}a_1a_{n+1}-qa_n&=(\alpha+\beta)(\alpha^{n+1}+\beta^{n+1})-q(\alpha^n+\beta^n)\\&=\alpha^{n+2}+\beta^{n+2}+\beta^n(\alpha\beta-q)+\alpha^n(\alpha\beta-q)\\&=\alpha^{n+2}+\beta^{n+2}\\&=a_{n+2},\end{aligned}$$

as needed. □

Exercise (Silverman 5.17). Let $E/\mathbb{F}_q$ be an elliptic curve and suppose that we know, a priori, that the zeta function of E has the form

$$Z(E/K;T)=\frac{1-aT+qT^2}{(1-T)(1-qT)}=\frac{(1-\alpha T)(1-\beta T)}{(1-T)(1-qT)},$$

with $a\in\mathbb{Z}$ and $\alpha,\beta\in\mathbb{C}$. Use this formula to prove that

$$\#E(\mathbb{F}_{q^n})=q^n+1-\alpha^n-\beta^n.\tag{4.2}$$

Proof. Taking the logarithm of the above equality of power series, and then expanding the four logarithms into power series, yields

$$\begin{aligned}\sum_{n\geq 1}\frac{\#E(\mathbb{F}_{q^n})}{n}T^n&=\log(1-\alpha T)+\log(1-\beta T)-\log(1-T)-\log(1-qT)\\&=\sum_{n\geq 1}\frac{-\alpha^n-\beta^n+1^n+q^n}{n}T^n,\end{aligned}$$

which immediately implies (4.2). □

5 Elliptic curves over local fields

5.1 Motivation, notation

A general number theory method for solving a problem:

1. Study the problem mod p.
2. Lift those solutions to study the problem in $\mathbb{Z}_p$ (this is Hensel's lemma).
3. Study the solutions in $\mathbb{R}$ (more generally, study the archimedian completions).
4. Fit together the $\mathbb{Z}_p$ and $\mathbb{R}$ information to study $\mathbb{Q}$.

We've done a fair amount studying elliptic curves over finite fields. So the next thing we want to do is study the elliptic curves over the p-adics. The following notation is in effect:

- K is a complete local field with respect to a normalized discrete valuation $v : K^* \twoheadrightarrow \mathbb{Z}$. Prototypical example is $\mathbb{Q}_p$.
- $R = \{\alpha \in K : v(\alpha) \geq 0\}$ is the ring of integers. Prototypical example is $\mathbb{Z}_p$.
- $R^* = \{\alpha \in K : v(\alpha) = 0\}$ is the unit group. Prototypical example is $\mathbb{Z}_p^*$.
- $\mathfrak{m} = \{\alpha \in K : v(\alpha) > 0\}$ is the maximal ideal. Prototypical example is $p\mathbb{Z}_p$.
- $\pi \in R$ is a uniformizer, i.e. $v(\pi) = 1$ and $\mathfrak{m} = \pi R$. Prototypical example is p.
- $k = R/\mathfrak{m}$ is the residue field. Prototypical example is $\mathbb{F}_p \cong \mathbb{Z}_p/pZ_p$.

We'll assume throughout that all these curves are perfect fields, and that $\operatorname{char}(k) \neq 2, 3$. *The overarching goal is to understand what the points on an elliptic curve look like on K by studying what they look like on R/π.*

5.2 Minimal Weierstrass equations

The idea: given an elliptic curve E/K with Weierstrass equation

$$E : y^2 = x^3 + Ax + B \qquad \text{where } A, B \in R,$$

we will reduce mod π to get some elliptic curve $\tilde{E}/k$ as well as a map $E(K) \to \tilde{E}(k)$. The problem: A and B might not reduce well modulo π, e.g. maybe $A, B \bmod \pi = 1/0$. To rescue the idea, we can change the coordinates of our Weierstrass equation; recall that we're allowed to use the variable transformations $x \mapsto u^{-2}x$ and $y \mapsto u^{-3}y$, which yields a new equation

$$y^2 = x^3 + u^4Ax + u^6B.$$

Choosing a value of u which is sufficiently divisible by π, we can clear all factors of π from the denominator; but if we choose a value of u which is too divisible by π, then we'll get $u^4A, u^6B \in \mathfrak{m}$, which will kill those terms, which is bad. We quantify this as follows. We will choose u so that:

1. The coefficients are in R (so we can reduce them modulo π)
2. $\Delta = 4A^3 + 27B^2$ is minimally divisible by π (so we didn't add too many factors of π)

With this goal in mind, the following definition is natural:

Definition 5.2.1. A *minimal Weierstrass equation* for E/K is one with:

- $E : y^2 = x^3 + Ax + B$, where $A, B \in R$.
- $v(\Delta(E))$ is minimized, subject to the requirement that $A, B \in R$.

Not too hard to show:

Fact 5.2.2. A Weierstrass equation is minimal if and only if $A, B \in R$, and $\min\{v(A^3), b(B^2)\} < 12$.

This condition is necessary because if the minimum is 12 or bigger, than we can find a u to get it down. The upshot of this? *Finding a minimum equation in the case where* $\operatorname{char} k \geq 5$ *is actually very easy.*

Now suppose $E : y^2 = x^3 + Ax + B$ is a minimal Weierstrass equation. This implies that A and B are determined up to a transformation $(A, B) \mapsto (u^4 A, u^6 B)$ with $u \in R^*$. Such a transformation maps the discriminant according to $\Delta \mapsto u^{12}\Delta$, which implies that $v(\Delta) = v(u^{12}\Delta)$. This implies that a minimal Weierstrass equation has a unique, well-defined valuation of the discriminant. We can also consider what happens to the invariant differential $\omega = dx/2y$. One can compute that $\omega \mapsto u\omega$ under this transformation, where u is a unit. The good thing about units is that you can reduce them mod π and still get a nonzero element of the finite field. In other words,

$$R^* \to k^* : u \mapsto \tilde{u} := u \pmod{\mathfrak{m}}$$

is a nice group homomorphism. Our next task is to apply use this group homomorphism to points of E.

(Lecture 21: March 10, 2021)

Given an elliptic curve $E : y^2 = x^3 + Ax + B$ with $A, B \in R$ and $v(\Delta)$ minimized, we can consider its reduction modulo π,

$$\tilde{E} : y^2 = x^3 + \tilde{A}x + \tilde{B}, \qquad \text{where } \tilde{A}, \tilde{B} \in R/\pi R = R/\mathfrak{m}.$$

So we an define the reduction mod π map

$$E \to \tilde{E} : (x, y) \mapsto \begin{cases} (\tilde{x}, \tilde{y}) & x, y \in R \\ \mathcal{O} & x \text{ or } y \notin R. \end{cases}$$

More generally, we define the reduction mod π map

$$\mathbb{P}^N(K) \to \mathbb{P}^N(k)$$

as follows: given $P = [a_0, \ldots, a_N] \in \mathbb{P}^N(K)$, so $a_i \in K$, we multiply these coordinates by some $u \in K^*$ to get $[b_0, \ldots, b_n]$ with all $b_i \in R$ and at least one $b_i \in R^*$. Then we define $\tilde{P} := [\tilde{b}_0, \ldots, \tilde{b}_n]$. Because one of the coordinates is a unit, one of the reduced coordinates is nonzero, so this is well-defined.

What happens when we reduce $E(K) \to \tilde{E}(k)$? We'd like for this to reduction be a homomorphism, but it's possible that $\tilde{E}(k)$ is singular, so it might not even be a group. It turns out:

Proposition 5.2.3. *If $\tilde{E}/k$ is non-singular,*[11] *then $E(K) \to \tilde{E}(k)$ is a surjective homomorphism.*

Proof. For surjectivity, let $f(x, y) = y^2 - x^3 - Ax - B$ be the generator for the ideal of the affine curve E. Let $(\tilde{\alpha}, \tilde{\beta}) \in \tilde{E}$. Recall Hensel's lemma: given a polynomial $F(T) \in R[T]$ and some $\tilde{\alpha} \in k$ such that $\tilde{F}(\tilde{\alpha}) = \tilde{0}$ and $\tilde{F}'(\tilde{\alpha}) \neq 0$, there exists $a \in R$ with $\tilde{a} = \tilde{\alpha}$ and $F(a) = 0$. In our case, we'll lift one coordinate arbitrarily, and find a corresponding second coordinate. Explicitly, we have $f(\tilde{\alpha}, \tilde{\beta}) = 0$. Because $\tilde{E}$ is non-singular, at least one of $\frac{\partial f}{\partial x}(\tilde{\alpha}, \tilde{\beta})$ and $\frac{\partial f}{\partial x}(\tilde{\alpha}, \tilde{\beta})$ is nonzero; we may assume $\frac{\partial f}{\partial x}(\tilde{\alpha}, \tilde{\beta}) \neq 0$. As $\tilde{\beta} \in k = R/\mathfrak{m}$, we can choose any $b \in R$ with $\tilde{b} = \tilde{\beta}$, and consider the polynomial $F(T) = f(T, b) \in R[T]$. Then $F'(T) = \frac{\partial f}{\partial x}(T, b)$, and we can compute that

$$F'(\tilde{\alpha}) = \frac{\partial \tilde{f}}{\partial x}(\tilde{\alpha}, \tilde{b}) = \frac{\partial \tilde{f}}{\partial x}(\tilde{\alpha}, \tilde{\beta}) \neq 0.$$

So by Hensel, there exists $a \in R$ with $\tilde{a} = \tilde{\alpha}$ and $F(a) = 0$, or equivalenrly, $f(a, b) = 0$. This proves that the reduction map is surjective.

It remains to show that the reduction map is a homomorphism. We will only argue one case in full detail; the remaining cases are in the textbook. Note that inverses clearly get sent to inverses, as the inverse of a point is obtained by negating the y coordinate, and this commutes with reduction modulo π. The idea of showing additivity is to show that if $P, Q, R \in E(K)$, then $P + Q + R = \mathcal{O}$ implies that $\tilde{P} + \tilde{Q} + \tilde{R} = \tilde{\mathcal{O}}$.

[11] This means $\tilde{\Delta} \neq 0$, or equivalently, $v(\Delta) = 0$.

- Case 1: $\tilde{P}, \tilde{Q}, \tilde{R}$ are distinct. In this case, we know $P + Q + R = \mathcal{O}$ if and only if P, Q, R are colinear, by definition of the group law. But this happens if and only if there exists a linear form $L(x, y, z) \in K[x, y, z]$ such that $L(P) = L(Q) = L(R) = 0$. This happens if and only if there exists $L \in R[x, y, z]$ with at least one coefficient in R^*. This implies $\tilde{L} = 0$ is a line in $\mathbb{P}^2(k)$, and $\tilde{L}(\tilde{P}) = \tilde{L}(\tilde{Q}) = \tilde{L}(\tilde{R}) = 0$. The assumption that P, Q, R have distint reductions means $\tilde{P}, \tilde{Q}, \tilde{R}$ are distinc, hence these three points are colinear in $\mathbb{P}^2(k)$. So by the definition of the group law on the reduced curve, this implies $\tilde{P} + \tilde{Q} + \tilde{R} = \tilde{\mathcal{O}}$ in $\tilde{E}(k)$, as claimed.

There are various other cases to check. □

In summary, we have a surjective group homomorphism $E(K) \to \tilde{E}(k) \to 0$. It is natural to ask what the kernel of the homomorphism is. If we unpack the definition, then we obtain the short exact sequence

$$0 \to E_1(K) \to E(K) \to \tilde{E}(k) \to 0,$$

where the kernel of the reduction map is

$$E_1(K) := \{P \in E(K) : \tilde{P} = \tilde{\mathcal{O}}\} = \{(x, y) \in E(K) : v(x), v(x) < 0\}.$$

And in fact, from the minimal Weierstrass equation $E : y^2 = x^3 + Ax + B$, we deduce that $A, B \in R$ implies $v(x) < 0$ if and only if $v(y) < 0$, and in this case, $3v(x) = 2v(y)$. This says the power of π in the denominator of x is a power of 2, and the power of π in the denominator of y is a power of 3. Something that we'll prove later:

Fact 5.2.4. $E_1(K)$ has a subgroup of finite index that is isomorphic to $(R, +)$. This isomorphism is not given by polynomials; it is given by convergent power series.

We want to discuss points of finite order in $E_1(K)$; to do this, we'll investigate the group $E_1(K)_{tors}$. Later we'll prove:

Theorem 5.2.5. *$E_1(K)$ has no m-torsion if* $\gcd(m, p) = 1$. *We say there is no "prime to p torsion," i.e. the only torsion would be p-power torsion.*

Let us discuss an analogy first, to understand where this important theorem comes from, and why it's important. Consider the multiplicative group

$$\mathbb{G}_m(R) := R^*.$$

In this case, we have the exact sequence

$$1 \to \{1\text{-units}\} \to R^* \to k^* \to 1,$$

where $\{1\text{-units}\} := \{u \in R^* : u \equiv 1 \pmod{\pi}\}$. In fancier notation, we write the sequence as

$$1 \to \mathbb{G}_m(R)_1 \to \mathbb{G}_m(R) \to \mathbb{G}_m(k) \to 1.$$

In this case, we understand $\mathbb{G}_m(k)$ really well, as this is a cyclic group of order $\#k - 1$. And the 1-unit group $\mathbb{G}_m(R)_1$ has a subgroup isomorphic to $(R, +)$. Note: we convert the additive group to the multiplicative group using logarithms.

Question: *What m-torsion can live in $\mathbb{G}(R)_1 = R_1^*$?* Here, a point $\zeta \in R$ is m-torsion means $\zeta^m = 1$.

Proposition 5.2.6. *Assume $p \nmid m$. Suppose $\zeta \in R$ satisfies $\zeta^m = 1$, and suppose $\tilde{\zeta} = \tilde{1}$, i.e. $\zeta \equiv 1 \pmod{\pi}$. Then $\zeta = 1$. In other words, the kernel of reduction doesn't contain any m'th roots of unity.*

Proof. As $\tilde{\zeta} = \tilde{1}$, we have $\zeta = 1 + \alpha\pi$ for some $\alpha \in R$. Thus

$$1 = \zeta^m = (1 + \alpha\pi)^m = 1 + m\alpha\pi + (\text{some mess}) \cdot \pi^2.$$

This implies that $m\alpha = (\text{some mess}) \cdot \pi$, so $m\alpha \in \pi R$. Because $p \nmid m$, p is a unit in R, so α is a multiple of π. So we really should have started with $\zeta = 1 + \alpha\pi^e$, where $\alpha \in R^*$ and $e > 1$. Re-doing the above arithmetic arithmetic yields

$$m\alpha = (\text{some mess}) \cdot \pi^e,$$

so $m\alpha \in \pi^e R$. This is a contradiction unless $\alpha = 0$, so $\zeta = 1$. □

The point of this proposition:

Corollary 5.2.7. *Assume $p \nmid m$ and $\zeta^m = 1$, where $\zeta \in \overline{K}$. Then $K(\zeta)/K$ is unramified.*

Proof. Assume towards a contadiction that $K(\zeta)/K$ is ramified. We know $K(\zeta)/K$ is ramified if and only if the minimal polynomial of ζ/K becomes inseparable (in other words, has a double root or more) when reduced modulo π. (Note: that's what ramification is; it's when you have things that are distinct, but which coincide once you reduce them.) But this implies that $x^m - 1 \pmod{\pi}$ has a double root. But this happens if and only if two of $1, \eta, \eta^2, \dots, \eta^{m-1}$ coincide modulo π, where η is a primitive m'th root of unity. But this means there are distinct mth roots of 1, say ζ_1, ζ_2, with $\zeta_1 \equiv \zeta_2 \pmod{\pi}$, or equivalently $\zeta_1\zeta_2^{-1} \equiv 1 \pmod{\pi}$. But $\zeta_1\zeta_2^{-1}$ is an m'th root of 1, so by the previous proposition, $\zeta_1\zeta_2^{-1} = 1$, so $\zeta_1 = \zeta_2$, which is a contradiction. □

Next time, we'll develop this machinery to do the same thing on elliptic curves.

(Lecture 22: March 12, 2021)

5.3 Motivation for formal groups

Given an elliptic curve E/K, we showed that if its reduction $\tilde{E}/k$ is nonsingular, then the reduction map $E(K) \to \tilde{E}(k)$ is surjective. This yields the short exact sequence

$$0 \to E_1(K) \to E(K) \to \tilde{E}(k) \to 0,$$

where by definition $E_1(K) = \{P \in E(K) : \tilde{P} = \tilde{\mathcal{O}}\}$, or equivalently, each $P = (x, y) \in E_1(K)$ satisfies $v(x), v(y) < 0$. So v-adically, this means $|x|_v, |y|_v > 1$. Let us apply a variable transformation to move $\mathcal{O} = [0, 1, 0]$ to the coordinates $(0, 0)$ in some affine patch. We take $z = -x/y$ and $w = -1/y$, so $x = z/w$ and $y = -1/w$, which means that

$$y^2 = x^3 + Ax + B \implies w = z^3 + Azw^2 + w^3.$$

If we substitute the expression for w on the RHS into each w on the RHS, and then do this again and again, then we'll get a power series

$$w = w(z) = z^3(1 + a_1 z + a_2 z^2 + \cdots) \in \mathbb{Z}[A, B][[z]]$$

which satisfies $f(z, w(z)) = 0$ in $R[[z]]$, where $f(z, w) = w - z^3 - Azw^2 - w^3$. Note that under this variable transformation, z is a local uniformizer at $\mathcal{O}$, i.e. $\mathrm{ord}_{\mathcal{O}}(z) = 1$. Geometrically, we should think of this as taking an infinitesimal neighborhood of the point $\mathcal{O}$. Furthermore, we can compute that

$$(z, w(z)) \in E_1(K)$$

as follows: we have

$$[x, y, 1] = [z/w, -1/w, 1] = [z, -1, w] = [z, -1, z^3 + \text{higher order terms}],$$

and when we reduce this modulo π, we get $\mathcal{O} = [0, 1, 0]$.

In summary, we've found a point $(z, w(z)) \in E_1(K)$, and the coordinates of this point are in a power series ring. Why is this useful? If we plug in any element of $\mathfrak{m} = \pi R$ into this power series ring, the power series will converge; because of the non-archimedian absolute value on $\mathbb{R}$, a power series converges if and only if the n'th term converges to 0. Thus, we obtain a map given by convergent power series.

$$\pi R \to E_1(K) : z \mapsto (z, w(z)).$$

In fact, this map will be a bijection (surjectivity can be proven using Hensel's lemma.) One can see that this power series map is continuous in the v-adic topology; so although this bijection doesn't describe $E_1(K)$ as a group, it describes it as a topological space, because of the topology in πR.

Next, we define the polynomial

$$f(z,w) = w - z^3 - Azw^3 - w^3.$$

We claim that there exists a unique $w(z) \in \mathbb{R}[[z]]$ such that $w(0) = 0$ and $f(z, w(z)) = 0$. But this follows from Hensel's lemma, because $\mathbb{R}[[z]]$ is a complete local ring (the valuation in any power series ring is just the number of z's that divide the power series). Furthermore, we can compute that

$$x(z) = \frac{z}{w(z)} = \frac{1}{z^2} \cdot (\text{something in } R[[z]]), \qquad y(z) = -\frac{1}{w(z)} = -\frac{1}{z^3} \cdot (\text{something in } R[[z]]),$$

and if we plug these into the invariant differential, then we can compute that

$$\omega(z) = \frac{dx(z)}{y(z)} = (\text{something in } R[[z]])dz.$$

At this point, we've defined a continuous bijection $\pi R \xrightarrow{\sim} E_1(K)$. As the target space $E_1(K)$ has a group law, our next task is to move this group law back to R. Under the map $\pi R \xrightarrow{\sim} E_1(K)$, we have z_1, z_2 get sent to $P_1 = (z_1, w(z_1)), P_2(z_2, w(z_2))$. The slope of the line through P_1 and P_2 is

$$\frac{w(z_2) - w(z_1)}{z_2 - z_1} = \frac{\text{some power series in } z_1, z_2 \text{ which vanishes at } z_1 = z_2}{z_2 - z_1} =: \lambda(z_1, z_2) \in R[[z_1, z_2]].$$

Then we can write the line through P_1, P_2 as

$$w = \lambda(z_1, z_2)z + \nu(z_1, z_2)$$

where again $\nu(z_1, z_2)$ will be some power series in $R[[z_1, z_2]]$. One can derive the formula for taking the inverse in the (z, w) plane and then obtain that the sum of z_1 and z_1 should be the power series $z_3(z_1, z_2) \in R[[z_1, z_2]]$ defined as

$$z_3(z_1, z_2) = -z_1 - z_2 - \frac{2A\lambda\nu + 3B\lambda^2\nu}{1 + A\lambda^2 + B\lambda^3} \in R[[z_1, z_2]],$$

or something like that; the details are in the book. The point is that λ, ν are power series in z_1, z_2 with no constant term, so using geometric series, the denominator has an inverse in the power series ring. This definition implies that z_3 is such that $P_1 + P_2 + P_3 = \mathcal{O}$, so $z(P_1 + P_2) = z_3(z_1, z_2)$. The upshot of this computation:

Proposition 5.3.1. *Given the map*

$$\pi R \xrightarrow{\sim} E_1(K) : z \mapsto (z, w(w))$$

defined above, there exists a power series $s(z_1, z_2) \in R[[z_1, z_2]]$ such that

$$(z_1, w(z_1)) + (z_2, w(z_2)) = (s(z_1, z_2), w(s(z_1, z_2))).$$

What have we accomplished? We have reduced addition using the group law in $E_1(K)$, which is given by rational functions, to imposing an addition law on the maximal ideal πR that is given by convergent power series. This is very useful in practice; by analogy, in complex analysis, the fact that we can expand a holomorphic function as a convergent power series is the starting point for the entire theory.

5.4 Formal groups

Informally, a formal group is a group law with no formal elements. More formally:

Definition 5.4.1. Let R be a ring. A *(one-parameter commutative) formal group $\mathscr{F}$ over R* is a power series $F(X, Y) \in R[[x, y]]$ satisfying:

1. $F(X, Y) = X + Y +$ (terms of degree ≥ 2), i.e. no constant terms.

2. $F(F(X,Y),z) = F(X,F(Y,Z))$, i.e. associativity.

3. $F(X,Y) = F(Y,X)$, i.e. commutativity.

4. There exists a unique $i(T) \in T \cdot R[[T]]$ such that $F(T,i(T)) = F(i(T),T) = 0$, i.e. inverses exist.

5. $F(X,0) = X$ and $F(0,Y) = Y$, i.e. there are no pure power terms other than X^1 and Y^1.

We should think of $F(X,Y)$ as "adding X and Y". It's an exercise that 1 and 2 imply 4 and 5; another exercise is that 1 and 2 the assumption that R has no nilpotent zero divisors implies 3.

Example 5.4.2.

1. The *formal additive group*, denoted $\hat{\mathbb{G}}_a$, is given by $F(X,Y) = X+Y$.

2. The *formal multiplicative group*, denoted $\hat{\mathbb{G}}_m$, is given by $F(X,Y) = (1+X)(1+Y)-1 = X+Y+XY$.

3. The *formal group of an elliptic curve*, denoted $\hat{E}$, is the power series $s(z_1,z_2)$ defined above. It tells you what addition looks like in the neighborhood of the identity element, in terms of formal power series.

Definition 5.4.3. Given formal groups $(\mathscr{F},F)$ and $(\mathscr{G},G)$, a *homomorphism* $\mathscr{F} \to \mathscr{G}$ is a power series $f(T) \in TR[[T]]$ satisfying

$$f(F(X,Y)) = G(f(x),f(y)).$$

We say $\mathscr{F}$ and $\mathscr{G}$ are *isomorphic* if there exist homomorphism $f : \mathscr{F} \to \mathscr{G}$ and $g : \mathscr{G} \to \mathscr{F}$ such that $f(f(T)) = T$ and $g(f(T)) = T$

Definition 5.4.4. The multiplication by m homomorphism $[m] : \mathscr{F} \to \mathscr{F}$ is defined inductively: $[0](T) = 0$, $[1](T) = T$, and $[m+1](T) = F([m](T),T)$. To define multiplication by a negative number, we define $[m-1](T) = F([m](T),i(T))$ and then use induction.

Can every group law be modeled this way? It turns out that the group law of any algebraic group over a complete local field can be modeled in a neighborhood of the identity. Conversely,

Fact 5.4.5. If R is a complete local ring with maximal ideal $\mathfrak{m}$, then a formal group $F \in R[[X,Y]]$ can be modeled as a commutative group

$$\mathscr{F}(\mathfrak{m}) = \begin{cases} \text{set:} & \mathfrak{m} \\ \text{operation:} & \alpha * \beta := F(\alpha,\beta). \end{cases}$$

Because R is complete, $F(\alpha,\beta)$ converges to some value of $\mathfrak{m}$ for every $\alpha,\beta \in \mathfrak{m}$. And because all the axioms of a formal group are satisfied for $\mathscr{F}$, we can deduce that all the axioms of a group are satisfied for $\mathscr{F}(\mathfrak{m})$.

Example 5.4.6. The kernel of the projection $R \to k$ is the group $\hat{\mathbb{G}}_a(R)$, i.e.,

$$0 \to \hat{\mathbb{G}}_a(R) \to R \to k \to 0,$$

and likewise

$$0 \to \hat{\mathbb{G}}_m(R) \to R^* \to k^* \to 0.$$

And in the exact sequence,

$$0 \to E_1(K) \to E(K) \to \tilde{E}(k) \to 1,$$

we have that there is a group isomorphism $E_1(K) \cong \hat{E}(\mathfrak{m})$.

This last example hopefully provides adequate motivation for studying formal groups.

(Lecture 23: March 15, 2021)

Lemma 5.4.7. *The multiplication-by-m homomorphism $[m] : \mathscr{F} \to \mathscr{F}$ satisfies the following properties:*

(a) $[m]T = mT +$ higher order terms.

(b) If $m \in R^*$, then $[m] : \mathscr{F} \to \mathscr{F}$ is an isomorphism.

Idea of proof. The first point follows from the fact that $F(X,Y) \in R[[X,Y]]$ has no constant term. The second point follows from Hensel's lemma; let $f(T) = [m](T) = mT +$ higher order terms. Construct $g(T) = m^{-1}T +$ higher order terms, and then use the assumption that $f(g(T)) = T$ to construct the coefficients of $g(T)$ inductively. □

Recall our local field setup $(K, R, \mathfrak{m}, k)$, and let $F(X,Y) \in R[[X,Y]]$ be a formal group over R. Then $\mathscr{F}(\mathfrak{m})$ is a group, with elements given by $\mathfrak{m}$, and the group law given by $\alpha * \beta = F(\alpha, \beta)$. Then we actually obtain a filtration

$$\mathscr{F}(\mathfrak{m}) \supseteq \mathscr{F}(\mathfrak{m}^2) \supseteq \mathscr{F}(\mathfrak{m}^3) \supseteq \dots.$$

This is because $\mathscr{F}(\mathfrak{m}^r)$ is closed under the group law; for if $\alpha, \beta \in \mathfrak{m}^r$, then $F(\alpha,\beta) = \alpha+\beta+$higher order terms, and every summand is in $\mathfrak{m}^r$. So we have a sequence of subgroups of an abelian group. As we can think of $\mathscr{F}(\mathfrak{m})$ as being built up from the successive quotient groups, it's valuable to study these quotients.

Proposition 5.4.8. *The map*

$$\mathscr{F}(\mathfrak{m}^n)/\mathscr{F}(\mathfrak{m}^{n+1}) \to \mathfrak{m}^n/\mathfrak{m}^{n+1}$$

induced by the identity map on sets is an isomorphism of groups. Here, the group law on the LHS is given by the formal group law; and the group law on the RHS is given by addition in $\mathbb{R}$.

Proof. For $\alpha, \beta \in \mathscr{F}(\mathfrak{m}^n)$, we have

$$\alpha * \beta = \mathscr{F}(\alpha,\beta) = \alpha + \beta + (\text{higher order terms in } \mathfrak{m}^{2n}) \equiv \alpha + \beta \pmod{\mathfrak{m}^{n+1}},$$

so this is indeed a homomorphism. □

The way to think about this result: both sets $\mathscr{F}(\mathfrak{m}^n)$ and $\mathfrak{m}^n$ are the same, as are the denominators $\mathscr{F}(\mathfrak{m}^{n+1})$ and $\mathfrak{m}^{n+1}$. So we're just looking at different group structures on these quotients; on the RHS we have regular addition, on the LHS we have the formal group law using power series.

Proposition 5.4.9. *Take the setup $(K, R, \mathfrak{m}, k)$ and let $p = \operatorname{char}(k)$. Let $\mathscr{F}/R$ be a formal group, and suppose $\alpha \in \mathscr{F}(\mathfrak{m})$ has finite order. Then its order is a power of p.*

Proof. We provide two proofs. For the first, let the order of α be $p^r m$ with $p \nmid m$. Then $\beta = [p^r](\alpha)$ has order m. We need to show that $\beta = 0$. Since $p \nmid m$, m is not in the maximal ideal, so it's a unit; but $m \in R^*$ implies that $[m] : \mathscr{F} \to \mathscr{F}$ is an isomorphism of the formal group. It follows that $[m] : \mathscr{F}(\mathfrak{m}) \to \mathscr{F}(\mathfrak{m})$ is an isomorphism of group associated to the formal group. So $[m](\beta) = 0$ implies $[m]^{-1}[m](\beta) = 0$, so $\beta = 0$.

For the second proof, let us assume β has order m and $p \nmid m$. We claim that $\beta \in \mathfrak{m}^r$ for all $r \geq 1$. For $r = 1$ this is clear, as we're only looking at points in the maximal ideal. Now assume $\beta \in \mathfrak{m}^r$, and $[m](\beta) = 0$. Consider the map

$$\mathscr{F}(\mathfrak{m}^r)/\mathscr{F}(\mathfrak{m}^{r+1}) \to \mathfrak{m}^r/\mathfrak{m}^{r+1}.$$

Then $[m](\beta) = 0$ on the LHS means $m\beta \equiv 0 \pmod{\mathfrak{m}^{r+1}}$ on the RHS. So $p \nmid m$ implies $m \notin \mathfrak{m}$, so $m \in \mathbb{R}^*$, therefore we can multiply this congruence by $m^{-1} \in R$ and obtain $\beta \in \mathfrak{m}^{r+1}$. We've now proven that

$$\beta \in \bigcap_{r \geq 1} \mathfrak{m}^r.$$

This intersection is $\{0\}$ because R is Noetherian, by Krull's theorem. □

Corollary 5.4.10. *Take the setup $(K, R, \mathfrak{m}, k)$ and let $p = \operatorname{char} k$. Let E/K be an elliptic curve, and assume that the reduction $\tilde{E}/k$ is smooth. In the exact sequence*

$$0 \to E_1(K) \to E(K) \to \tilde{E}(k) \to 0,$$

we have that

$$E_1(K) \cong \hat{E}(\mathfrak{m}),$$

where $\hat{E}$ is the formal group associated to the elliptic curve E. Furthermore, $\hat{E}(\mathfrak{m})$ has no prime-to-p torsion.

In other words, if we have m-torsion points in $E(K)$, then they'll inject into $\tilde{E}(k)$ when we reduce modulo p. This is a useful property to have at our disposal, because reduction modulo $\mathfrak{m}$ is actually the way to study ramification. More concretely, if $0 \neq P \in E(K)[m]$ and $m \not\equiv 0 \pmod{p}$, then $\tilde{P} \neq \tilde{\mathcal{O}}$. To restate this in terms of kernels: if $p \nmid m$ (equivalently, if $\tilde{m} \neq \tilde{0}$ in k) then the m torsion of the elliptic curve, $E(K)[m]$, injects into the m-torsion on the reduced curve, $\tilde{E}(k)[m]$. This is analogous to the theorem that we proved for cyclotomic fields, that says: if $p \nmid m$, then the set of m'th roots of unity in the local field K injects into the m'th roots of unity in k^*.

Corollary 5.4.11. *Let $K/\mathbb{Q}$ be a finite extension, and let E/K an elliptic curve. Then $E(K)_{tors}$ is finite.*

Proof. Let $\mathfrak{p}, \mathfrak{q}$ be primes of K where

(a) E has good reduction at $\mathfrak{p}$ and $\mathfrak{q}$, and

(b) If $p = \operatorname{char}(k_{\mathfrak{p}})$ and $q = \operatorname{char}(k_{\mathfrak{q}})$, then $p \neq q$.

Denote by $K_{\mathfrak{p}}$ the local field which is the completion of K at $\mathfrak{p}$. Then $K \subseteq K_{\mathfrak{p}}$, so clearly

$$E(K)[m] \subseteq E(K_{\mathfrak{p}})[m] \qquad \implies \qquad E(K)_{\text{prime-to-}p\text{-torsion}} \subseteq E(K_{\mathfrak{p}})_{\text{prime-to-}p\text{-torsion}}.$$

But by the above corollary, if $(m, p) = 1$, then there is the inclusion $E(K_{\mathfrak{p}})[m] \hookrightarrow \tilde{E}_{\mathfrak{p}}(k_{\mathfrak{p}})[m]$. Composing these inclusions yields the injections

$$E(K)_{\text{prime-to-}p\text{-torsion}} \hookrightarrow \tilde{E}_{\mathfrak{p}}(k_{\mathfrak{p}}), \qquad E(K)_{\text{prime-to-}q\text{-torsion}} \hookrightarrow \tilde{E}_{\mathfrak{q}}(k_{\mathfrak{q}}).$$

But all of the torsion in $E(K)$ lives in one of these two sets on the LHS; thus, we can estimate that

$$\begin{aligned} \#E(K)_{tors} &\leq (\#\text{prime-to-}p\text{-torsion points})(\#\text{prime-to-}p\text{-torsion points}) \\ &\leq (\#\tilde{E}(\mathbb{F}_p))(\#\tilde{E}(\mathbb{F}_q)) \\ &\leq (\sqrt{N\mathfrak{p}} + 1)^2(\sqrt{N\mathfrak{q}} + 1)^2, \end{aligned}$$

where the last bound is Hasse's theorem. □

A much stronger theorem:

Theorem 5.4.12 (Uniformity of torsion on elliptic curves)**.** *For all integers $d \geq 1$, there exists a bound $B(d)$ such that for all number fields $K/\mathbb{Q}$ with $[K : \mathbb{Q}] \leq d$, for all elliptic curve E/K, we have*

$$\#E(K)_{tors} \leq B(d).$$

The proof of this for $d = 1$ is due to Mazur, who showed that $B(1) = 16$; for $d = 2, \ldots, 8$, it's due to Kamienny; various people (including Dan Abramovich) extended Kamienny's argument; and the case of general d was proven by Merel. The hard case of this theorem is if E has bad reduction at all small primes. A fun fact: if you know the *abc*-conjecture, then you can prove Merel's theorem fairly easily, because the *abc*-conjecture precludes the possibility of bad reduction at all small primes.

(Lecture 24: March 17, 2021)

Today we'll talk a bit more about formal groups. Our goal today is to show that $\mathscr{F}(\mathfrak{m})$ contains a big subgroup that looks like the additive group $(R, +)$.

Definition 5.4.13. An *invariant differential* on a formal group $\mathscr{F}/R$ is some

$$\omega(T) = P(T)dT \in R[[T]]\,dT$$

which satisfies $\omega \circ F(T, S) = \omega(T)$. Informally speaking, this means ω is invariant under translation by S.

We can compute that the above condition is equivalent to

$$P(F(T, S))\frac{\partial F}{\partial X}(T, S) = P(T).$$

Example 5.4.14.

- In the additive group $\hat{\mathbb{G}}_a$, an invariant differential is $\omega(T) = dT$.
- In the multiplicative group $\hat{\mathbb{G}}_m$, an invariant differential is $\omega(T) = \frac{dT}{1+T}$.

Proposition 5.4.15. *For any formal group $\mathscr{F}/R$, there exists a nonzero invariant differential. Furthermore, an invariant differential is unique up to multiplication by $c \in R$. In particular, there is a unique invariant differential of the form*

$$\omega(T) = (1 + \textit{higher order terms})dT.$$

Proof. A differential $\omega(T) = P(T)dT$ is invariant if and only $P(F(T,S))F_X(T,S) = P(T)$ in $R[[T,S]]$. In particular, this identity must hold when $T = 0$, so this implies $P(F(0,S))F_X(0,S) = P(0)$. But remember that $F(0,S) = S$ and $P(0) \in R$, and $F_X(0,S) = 1 +$ higher order terms. This implies that, if there is an invariant differential, then it has to be of the form

$$\omega(T) = P(0) \cdot \frac{dT}{F_X(0,T)}.$$

This proves uniqueness up to multiplication by a constant.

For existence, let $\omega(T) = dT/F_X(0,T) = (1 + \text{higher order terms})dT$. We claim that $\omega(T)$ is invariant. To prove this, it suffices to show that

$$F_X(0, F(T,S))^{-1} F_X(T,S) = F_X(0,T)^{-1}.$$

To prove this, we apply the chain rule to the associative law: the associative law says $F(U, F(T,S)) = F(F(U,T),S)$, and if we take $\frac{\partial}{\partial U}$ and use the chain rule, then we get

$$F_X(U, F(T,S)) = F_X(F(U,T),S)F_X(U,T)$$

for all U, S, T. So take $U = 0$. □

Now take any homomorphism $f : \mathscr{F} \to \mathscr{G}$, so $f(T) = aT + bT^2 + cT^3 + \cdots \in TR[[T]]$. One can compute directly that $f^*\omega_{\mathscr{G}} = \omega_{\mathscr{G}} \circ f$ is an invariant differential on $\mathscr{F}$, so it's equal to some constant times $\omega_{\mathscr{F}}$. But the constant term on $\omega_{\mathscr{G}} \circ f$ is a, and the constant term on $\omega_{\mathscr{F}}$ is 1, so we've shown that

$$\omega_{\mathscr{G}} \circ f = f'(0)\omega_{\mathscr{F}}.$$

Now consider a formal group $\mathscr{F}$ and a prime p. We want to compute the multiplication by p map in the formal group; denote this map by $[p]_{\mathscr{F}}(T)$. We know $[p]_{\mathscr{F}}(T) = pT +$ higher order terms. This implies that $[p]'_{\mathscr{F}}(0) = p$. So applying the above formula, we obtain

$$p(1 + \text{higher order terms}) = p\omega(T) = (\omega \circ [p]_{\mathscr{F}})(T) = (1 + \text{higher order terms})[p]'(T)dT,$$

where we wrote $\omega(T) = P(T)dT$. It follows that every coefficient of $[p]'(T)$ is a multiple of p. This tells us that $[p](T) = \sum c_n T^n$, where $nc_n = p \cdot a_n$. We can gather the exponents (resp. the coefficients) which are divisible by p. In summary, we've shown:

Proposition 5.4.16. *There exist $g(T), h(T) \in TR[[T]]$ such that the multiplication by p map in $\mathscr{F}$ has the form*

$$[p]_{\mathscr{F}}(T) = pg(T) + h(T^p).$$

Next, we recall that the differential linearizes the group law; so, in order to get from the differential to the additive group, we'll integrate.

Definition 5.4.17. Let $\omega_{\mathscr{F}}(T)$ be the normalized invariant differential. The *formal logarithm* on $\mathscr{F}$ is

$$\log_{\mathscr{F}}(T) = \int \omega_{\mathscr{F}}(T).$$

If the invariant differential is $\omega_{\mathscr{F}}(T) = (1 + c_1 T + c_2 T^2 + \cdots)dT$, then we can compute that

$$\int \omega_{\mathscr{F}}(T) = T + \frac{c_1}{2}T^2 + \frac{c^2}{3}T^3 + \cdots \in K\left[[T]\right] dT,$$

where we assume that $R \hookrightarrow K := \mathbb{R} \otimes \mathbb{Q}$, because the coefficients here might not be in R anymore. We need this injectivity assumption because bad things might happen in characteristic p. Next, note that the leading term of the logarithm is T, and 1 is a unit in any ring, so this has a formal inverse:

Definition 5.4.18. The *formal exponential* is the power series $\exp_{\mathscr{F}}(T) \in K\left[[T]\right]$ which satisfies

$$\log_{\mathscr{F}} \circ \exp_{\mathscr{F}}(T) = \exp_{\mathscr{F}} \circ \log_{\mathscr{F}}(T) = T.$$

Lemma 5.4.19. *Consider the power series $f(T) = \sum_{n=1}^{\infty} \frac{a_n}{n!}T^n$ with $a_1 = 1$ and every $a_n \in R$. Then, the unique $g(T)$ with $f \circ g(T) = g \circ f(T) = T$ is of the form*

$$g(T) = \sum_{n=1}^{\infty} \frac{b_n}{n!}T^n,$$

where every $b_i \in R$.

Informally speaking, this says that the coefficients of f^{-1} are no worse than the coefficients of f.

Corollary 5.4.20. *$\log_{\mathscr{F}}(T) = \sum_{n=1}^{\infty} \frac{c_n}{n}T^n$, where $c_n \in R$, and $\exp_{\mathscr{F}}(T) = \sum_{n=1}^{\infty} \frac{b_n}{n!}T^n$, where $b_n \in R$.*

Corollary 5.4.21. *$\log_{\mathscr{F}} : \mathscr{F} \to \hat{\mathbb{G}}_a$ is an isomorphism over $K = R \otimes \mathbb{Q}$.*

Proof. $\exp_{\mathscr{F}}$ is the inverse. □

A priori, the previous corollary only gives an isomorphism of *formal* groups; but this implies that there is an isomorphism of groups if we're working over a ring where there is convergence. So take our local ring setup $(K/\mathbb{Q}_p, R, \mathfrak{p}, k = R/\mathfrak{p})$, let π be a uniformizer and let $v : K^* \to \mathbb{Z}$ be a normalized valuation, so $\mathfrak{p} = \pi R$ where $v(\pi) = 1$. Then the ramification index is $v(p) = e_{\mathfrak{p}}(K/\mathbb{Q}_{\mathfrak{p}})$ and $pR = \pi^{v(p)}R$. Exercises:

- $\frac{v(n)}{n} \to 0$ as $n \to \infty$. This follows from the fact that $v(n) \leq \log_p(n)$.
- $v(n!) \leq \frac{(n-1)v(p)}{p-1}$. The idea of proving this:

$$v(n!) = \sum_{k=1}^{\infty} \#\{1 \leq j \leq n : p^k \mid j\} v(p) = \sum_{k=1}^{\infty} \left\lfloor n/p^k \right\rfloor v(p),$$

  and if we take the floors out, we'll get something a tad worse than the claimed upper bound; and by doing it carefully, we'll get the better bound.

Corollary 5.4.22. *There is a well-defined injective group homomorphism*

$$\mathscr{F}(\mathfrak{p}) \hookrightarrow \hat{\mathbb{G}}_a(\mathfrak{p}).$$

Proof. We know $\log_{\mathscr{F}}(T) = \sum_{n=1}^{\infty}(c_n/n)T^n$ where each $c_n \in R$. If $t \in \mathfrak{p}$, then

$$v(c_n t^n/n) = v(c_n) + nv(t) - v(n) \geq nv(t) - v(n) \geq n - \log_p(n) \to \infty$$

as $n \to \infty$, which means $|c_n t^n/n|_{\mathfrak{p}} \to 0$, so the series $\log_{\mathscr{F}}(T)$ converges to an element of $\mathfrak{p}$. □

Note that $\exp_{\mathscr{F}} : \hat{\mathbb{G}}_a(\mathfrak{m}) \to \mathscr{F}(\mathfrak{m})$ gives an inverse if this converges. In general this won't converge on the maximal ideal (because of all the factorials in the denominator) but it will comverge on some power of the maximal ideal. If $\exp_{\mathscr{F}} : \hat{\mathbb{G}}_a(\mathfrak{m}^r) \to \mathscr{F}(\mathfrak{m}^r)$ does convverge for some $t \in \mathfrak{m}^r$, then the valuation of the n'th term of the series at some $t \in \mathfrak{p}$ is

$$v(b_n t^n/n!) = v(b_n) + nv(t) - v(n!) \geq nv(t) - \frac{nv(p)}{p-1} = n\left(v(t) - \frac{v(p)}{p-1}\right) = n\left(r - \frac{v(p)}{p-1}\right) \to \infty$$

as $n \to \infty$, so long as $r > v(p)/(p-1)$. What we've proven:

Corollary 5.4.23. *Assume $r > \frac{v(p)}{p-1}$. Then in the short exact sequence*

$$0 \to \mathscr{F}(\mathfrak{m}^r) \to \mathscr{F}(\mathfrak{m}) \to \mathscr{F}(\mathfrak{m})/\mathscr{F}(\mathfrak{m}^r) \to 0,$$

we have $\mathscr{F}(\mathfrak{m}^r) \cong (\mathfrak{m}^r, +)$, and $\mathscr{F}(\mathfrak{m})/\mathscr{F}(\mathfrak{m}^r)$ is a finite additive group of order $\#(\mathfrak{m}/\mathfrak{m}^r) = \#k^{r-1}$.

The upshot of this corollary, and the point of today's lecture? We've broken up $\mathscr{F}(\mathfrak{m})$ into an additive group, and a finite group whose order we understand. The finite group will bound the torsion in the elliptic curve; and in the cases where $r = 1$ suffices, the quotient $\mathscr{F}(\mathfrak{m})/\mathscr{F}(\mathfrak{m}^r)$ is zero, so we can in fact conclude that $(\mathfrak{m}^r, +) \cong \mathscr{F}(\mathfrak{m})$, i.e., there is no torsion.

5.5 Exercises

Exercise (Silverman 7.4). Let E/K be an elliptic curve given by a minimal Weierstrass equation, and for each $n \geq 1$, define a subset of $E(K)$ by

$$E_n(K) := \{P \in E(K) : v(x(P)) \leq -2n\} \cup \{\mathcal{O}\}.$$

(a) Prove that $E_n(K)$ is a subgroup of $E(K)$.

(b) Prove that

$$E_n(K)/E_{n+1}(K) \cong k^+.$$

Proof. Let $E : y^2 = x^3 + Ax + B$ be a minimal Weierstrass equation for E/K. Recall that we have the short exact sequence

$$0 \to E_1(K)(\cong \hat{E}(\mathfrak{m})) \to E(K) \to \hat{E}(k) \to 0,$$

where $\hat{E}$ is the formal group associated to E. This short exact sequence implies that $\hat{E}(\mathfrak{m})$ embeds into $E(K)$, so for (a) it suffices to show that $E_n(K)$ is isomorphic to a subgroup of $\hat{E}(\mathfrak{m}^n)$, as $\hat{E}(\mathfrak{m}^n)$ is a subgroup of $\hat{E}(\mathfrak{m})$. Towards this, recall that we constructed an isomorphism

$$\hat{E}(\mathfrak{m}) \to E_1(K) : z \mapsto \left(\frac{z}{w(z)}, -\frac{1}{w(z)}\right).$$

But in fact, this gives rise to an isomorphism

$$\hat{E}(\mathfrak{m}^n) \to E_n(K) : z \mapsto \left(\frac{z}{w(z)}, -\frac{1}{w(z)}\right),$$

because $z \in \mathfrak{m}^n$ implies $v(z/w(z)) = -2v(z) \leq -2n$. And (b) is true because (a) implies that

$$E_n(K)/E_{n+1}(K) \cong \hat{E}(\mathfrak{m}^i)/\hat{E}(\mathfrak{m}^{i+1}) \cong \mathfrak{m}^i/\mathfrak{m}^{i+1},$$

and the latter quotient is a one-dimensional k-vector space, so $\mathfrak{m}^i/\mathfrak{m}^{i+1} \cong (k, +)$. □

(Lecture 25: March 19, 2021)

6 Elliptic curves over global fields

6.1 Weak Mordell-Weil theorem

Our short term goal is to prove the following:

Theorem 6.1.1 (Mordell-Weil theorem)**.** *Let $K/\mathbb{Q}$ be a number field and E/K an elliptic curve. Then $E(K)$ is a finitely generated abelian group.*

We call $E(K)$ the *Mordell-Weil group* of the elliptic curve. The proof of this result naturally falls into two pieces. We'll first show:

Theorem 6.1.2 (Weak Mordell-Weil theorem)**.** *For all $m \geq 2$, $E(K)/mE(K)$ is finite.*

The Mordell-Weil theorem implies the weak Mordell-Weil theorem, but the converse is not immediate. For example, note that $\mathbb{Q}/m\mathbb{Q} = 0$ for all m, but $\mathbb{Q}$ is not finitely generated as an abelian group; so there is some work to do to go from the weak to the full theorem.

Proposition 6.1.3. *Let $K/\mathbb{Q}$ be a number field and E/K an elliptic curve. Let $\mathfrak{p} \subseteq K$ be a prime and $m \geq 2$ an integer. Assume that $\mathfrak{p} \nmid m$ and that E has good reduction at $\mathfrak{p}$. Then $E[m](K) \to \tilde{E}_{\mathfrak{p}}(k_{\mathfrak{p}})$ is injective.*

Corollary 6.1.4. *Let $K/\mathbb{Q}$ be a number field,let E/K be an elliptic curve with good reduction at $\mathfrak{p}$, and assume $\mathfrak{p} \nmid m$. Then the extension $K(E[m])/K$ is unramified over $\mathfrak{p}$.*

Proof. Let $\mathfrak{P}$ be a prime ideal in the ring of integers of $K(E[m])$ which lies over $\mathfrak{p}$. We can write $K(E[m]) = K(T_1, \ldots, T_{m^2})$, where the T_i are the coordinates of the m-torsion points. If $\mathfrak{P}/\mathfrak{p}$ is ramified, then two of the T_i coincide when reduced mod $\mathfrak{p}$. (In other words, if none of the generators come together when reduced $\mathfrak{p}$, then $\mathfrak{P}/\mathfrak{p}$ is unramified, because the roots of the minimal polynomial stay separated.) This implies that $\tilde{T}_i = \tilde{T}_j \pmod{\mathfrak{p}}$ in $\tilde{E}_{\mathfrak{p}}(k_{\mathfrak{p}})$, which in turn implies that

$$T_i - T_j \in \ker\big(E[m](K(E[m])) \to \tilde{E}_{\mathfrak{p}}(k_{\mathfrak{p}})\big).$$

But the theorem we just proved says that $E[m](K(E[m])) \to \tilde{E}_{\mathfrak{p}}(k_{\mathfrak{p}})$ is injective (good reduction at $\mathfrak{p}$ implies good reduction at $\mathfrak{P}$, and $\mathfrak{p} \nmid m$ implies $\mathfrak{P} \nmid m$) which implies that $T_i = T_j$. This is a contradiction, thus $\mathfrak{P}/\mathfrak{p}$ is unramified. □

Proof of Weak Mordell-Weil theorem. First, some motivation. How would we prove that $E(K)/mE(K)$ is finite? We want to study the following question: given $P \in E(K)$, to what extent is $P = mQ$ for some $Q \in E(K)$? I.e., among m^2 different solutions to $mQ = P$ with $Q \in E(\overline{K})$, are there any in $E(K)$? If there are, then $P = 0$ in $E(K)/mE(K)$, and if there aren't, then $P \neq 0$ in $E(K)/mE(K)$. So, how do we check if an algebraic number in $\overline{K}$ is in the base field? We use Galois theory. Let $Q \in E(\overline{K})$ with $mQ = P \in E(K)$. Then $Q \in E(K)$ if and only if $Q^\sigma = Q$ for all $\sigma \in G_K := \mathrm{Gal}(\overline{K}/K)$. But this is true if and only if $Q^\sigma - Q = \mathcal{O}$. So it's natural to define the map

$$G_K \to E(\overline{K}) : \sigma \mapsto Q^\sigma - Q.$$

To some extent, this map is a measure of how big a field Q generates; if it's the zero map, then $Q \in E(K)$, i.e., Q lives in the base field. Because the group law commutes with the action of Galois, we can observe that $m(Q^\sigma - Q) = (mQ)^\sigma - mQ = P^\sigma - P = \mathcal{O}$. This tells us that the image points are actually m-torsion points, so in fact we have a map

$$G_K \to E(\overline{K})[m] : \sigma \mapsto Q^\sigma - Q.$$

In general this is not a group homomorphism; in fact, it's a group cocyle. We want to reduce to the case where this is a group homomorphism.:

Reduction step: Let K'/K be a finite extension. Then $E(K)/mE(K)$ is finite if and only if $E(K')/mE(K')$ is finite. What this means is that, in proving the weak Mordell-Weil theorem, we can choose to work over an extension field. So without loss of generality, we may assume $E[m] \subseteq E(K)$, i.e., we can replace K with the field where we adjoin all the K-torsion points.

Claim 6.1.5. Let $Q \in E(\overline{K})$ such that $mQ = P \in E(K)$. Then the map

$$\kappa_Q : G_K \to E[m] : \kappa_Q(\sigma) = Q^\sigma - Q$$

is a group homomorphism.

Proof of claim. We want to show that $\kappa(\sigma\tau) = \kappa(\sigma)\kappa(\tau)$. We can compute

$$\kappa(\sigma\tau) = Q^{\sigma\tau} - Q = Q^{\sigma\tau} - Q^\tau + Q^\tau - Q = (Q^\sigma - Q)^\tau + (Q^\tau - Q) = \kappa(\sigma)^\tau + \kappa^\tau.$$

But $\kappa(\sigma)^\tau = \kappa(\sigma)$ because $\kappa(\sigma) \in E[m]$, and we assumed $E[m] \subseteq E(K)$, i.e. the m-torsion points are fixed by Galois. □

Claim 6.1.6. $\kappa_Q(\sigma)$ only depends on $P = mQ$.

Proof of claim. Assume $P = mQ = mQ'$. Then $m(Q - Q') = \mathcal{O}$, so $Q - Q' \in E[m]$. Then we can compute

$$\kappa_Q(\sigma) - \kappa_{Q'}(\sigma) = (Q^\sigma - Q) - (Q'^\sigma - Q') = (Q - Q')^\sigma - (Q - Q').$$

But $Q - Q' \in E[m] \subseteq E(K)$, so $Q - Q'$ is fixed by σ, which implies that $\kappa_Q(\sigma) - \kappa_{Q'}(\sigma) = \mathcal{O}$. □

In summary, the following is a well-defined group homomorphism:

$$\kappa : E(K) \to \mathrm{Hom}(G_K, E[m]) : P \mapsto \kappa_Q \text{ for any } Q \in [m]^{-1}(P).$$

By abuse of notation we'll write $\kappa_P := \kappa_Q$ because it only depends on P.

Claim 6.1.7. $mE(K)$ is the kernel of this homomorphism.

Proof of claim. We have that

$$\begin{aligned}
\kappa_P = 0 &\iff \kappa_P(\sigma) = 0 \text{ for all } \sigma \in G_K \\
&\iff Q^\sigma - Q = 0 \text{ for all } \sigma \in G_K, \text{ where } Q \in [m]^{-1}(P) \\
&\iff Q \in E(K) \\
&\iff P = mQ \in mE(K).
\end{aligned}$$

□

Modding out by this kernel gives an injective group homomorphism

$$\kappa : E(K)/mE(K) \hookrightarrow \mathrm{Hom}(E_K, E[m]) : \kappa_P(\sigma) = Q^\sigma - Q \text{ for any } Q \in [m]^{-1}(P).$$

The torsion group $E[m]$ is finite of size m^2, but G_K is a huge profinite group. Our goal is to restrict G_K to some finite subgroup, because there are only finitely many homomorphisms between finite groups. Towards this, consider the field extension

$$L := K(Q : mQ \in E[K]).$$

Note that if $\sigma \in \mathrm{Gal}(\overline{K}/L)$, then $\kappa_P(\sigma) = Q^\sigma - Q = \mathcal{O}$, which implies that the κ_P kill the Galois group $\mathrm{Gal}(\overline{K}/L)$. So the key proposition that will allow us to almost complete the proof of the weak Mordell-Weil theorem is the following:

Proposition 6.1.8. *There is an injectiion*

$$\kappa : E(K)/mE(K) \hookrightarrow \mathrm{Hom}(G_{L/K}, E[m]),$$

where $L := K(Q : [m]Q \in E[K])$.

Note that once we prove this proposition, to show the weak Mordell-Weil theorem, it will suffice to show that L/K is finite.

Proof of proposition. By the previous claim, we have a pairing

$$G_{L/K} \times E(K)/mE(K) \to E[m] : (\sigma, P) \mapsto Q^\sigma - Q \text{ for any } Q \in [m]^{-1}P.$$

But in fact this is a perfect pairing: if we fix σ and get $(\sigma, P) \mapsto \mathcal{O}$ for all P, then in fact $\sigma = 1$, i.e. σ fixes L, because $Q_P^\sigma - Q_P = \mathcal{O}$ for all P implies σ fixes every Q satisfying $mQ \in E(K)$, and those are exactly the Q's which generate L. □

Comversely, we get injective homomorphism

$$\psi : G_{L/K} \hookrightarrow \mathrm{Hom}(E(K)/mE(K), E[m]) : \sigma \mapsto (P \mapsto \kappa_P(\sigma)),$$

which is just the other half of the perfect pairing. To recapitulate, $G_{L/K}$ is abelian and has exponent dividing m, i.e. $\sigma^m = 1$ for all $\sigma \in G_{L/K}$. These conditions don't in general guarantee that L/K is finite; for example, $\mathbb{Q}(\sqrt{p} : p)/\mathbb{Q}$ is abelian, and is a 2-group (every element has order 1 or 2.) So next time we'll add in the one extra ingredient to show that L/K is finite.

Our goal is to show that $G_{L/K}$ is finite. Note that the injective homomorphism ψ implies that $G_{L/K}$ is abelian, since it's contained in an abelian group; and it has exponent dividing m as we computed above; furthermore we claim that

$$\#\{\mathfrak{p} : L/K \text{ is ramified at } \mathfrak{p}\} < \infty. \tag{6.1}$$

Then we'll show that these three properties, if they're true for any number field, imply that L/K is finite. (Note that if L/K were finite, it would be ramified at finitely many primes; so we're after a sort of converse.)

Proof of (6.1). We will first throw away finitely many primes: we will assume E has good reduction at $\mathfrak{p}$, and we'll assume $\mathfrak{p} \nmid m$. If we assume that L/K is ramified at $\mathfrak{p}$, then we'll arrive at a contradiction, as this will show that the set of primes in (6.1) contains at most those primes where E has bad reduction at $\mathfrak{p}$ and where $\mathfrak{p} \mid m$. As L/K is ramified at $\mathfrak{p}$, there exists a finite subextenstion $K \subseteq L' \subseteq L$ such that there is some ramified prime $\mathfrak{P} \subseteq L'$ lying above $\mathfrak{p} \subseteq K$. If L' contains some point Q such that $[m]Q \in E[K]$, let us assume L' contains all Galois conjugates of all Q as well; then L'/K is still finite. But recall that we defined $L = \{Q \in E(\overline{K}) : [m]Q \in E(K)\}$. So the fact that $\mathfrak{P}/\mathfrak{p}$ is ramified means there exists $\sigma \in G_{L/K}$ so that $\sigma(Q) \neq Q$ but $\widetilde{\sigma(Q)} \equiv \tilde{Q} \pmod{\mathfrak{P}}$ (by definition, a prime is ramified if and only if the corresponding inertia group is nontrivial; the elements of the inertia group fix everything $\bmod \mathfrak{P}$, so the inertia group is nontrivial if and only if an element is nontrivial but fixes things modulo $\mathfrak{P}$.) Now let $T = \sigma(Q) - Q$. Then $T \in E[m]$, and we know

$$\tilde{T} = \widetilde{\sigma(Q) - Q} = \widetilde{\sigma(Q)} - \tilde{Q} = \tilde{\mathcal{O}}.$$

But this contradicts the fact that $E[m]$ injects into $\tilde{E}(k_{\mathfrak{p}})$ (we showed the map $E[m] \to \tilde{E}(k_{\mathfrak{p}})$ is an injection when we reduce modulo a prime of good reduction. □

Now it remains to show the following statement from algebraic number theory:

Theorem 6.1.9. *Let $K/\mathbb{Q}$ be a number field, $m \geq 2$, S a finite set of primes of K. Let L be the maximal extension of K satisfying:*

1. *L/K is abelian.*
2. *$G_{L/K}$ has exponent m.*
3. *L/K is unramified outside of S.*

Then L/K is a finite extension.

Example 6.1.10. Consider the special case $K = \mathbb{Q}, m = 2, S = \{2, p_1, \dots, p_r\}$. Let L be as above. Then it's a relatively easy exercise that $L = \mathbb{Q}(\sqrt{2}, \sqrt{p_1}, \dots, \sqrt{p_r}, \sqrt{-1})$, so

$$G_{L/K} \cong (\mathbb{Z}/2\mathbb{Z})^{r+2} = (\mathbb{Z}/2\mathbb{Z})^{\#S+1}.$$

Let's translate this back to the elliptic curve case. Suppose we take an elliptic curve with the 2-torsion rational, so

$$E : y^2 = (x-a)(x-b)(x-c), \qquad a, b, c \in \mathbb{Z}.$$

Then $E[2] = \{\mathcal{O}, (a,0), (b,0), (c,0)\} \subseteq E(\mathbb{Q})$. So the primes of bad reduction is the set $\{p : p \mid \Delta_E\}$; this is because a prime divides the discriminant means the elliptic curve is singular when reduced modulo that prime, which means the cubic on the RHS gets a double root, which means two of the roots become congurent when reduced by p. In other words, the set of primes of bad reduction is exactly

$$\{p \mid (a-b)(a-c)(b-c) \cdot 2\}.$$

Write $\nu(E)$ to be the number of primes dividing Δ_E. Then we will have proved that

$$E(\mathbb{Q})/2E(\mathbb{Q}) \hookrightarrow \operatorname{Hom}(G_{L/\mathbb{Q}}, E[2]).$$

But $G_{L/\mathbb{Q}} \subseteq (\mathbb{Z}/2\mathbb{Z})^{\nu(\Delta_E)+1}$, and $E[2] \cong (\mathbb{Z}/2\mathbb{Z})^2$. So we get that not only $E(\mathbb{Q})/2E(\mathbb{Q})$ is a finite group, but also we get an explicit bound

$$\#(E(\mathbb{Q})/2E(\mathbb{Q})) \leq 2^{2(\nu(\Delta_E)+1)}.$$

Mordell-Weil implies that

$$E(\mathbb{Q}) \cong E[2] \times \text{ a finite group of odd order } \times \mathbb{Z}^r,$$

so modding out by $2E(\mathbb{Q})$, we get that the rank is in fact

$$r = \operatorname{rank} E(\mathbb{Q}) \leq 2\nu(\Delta_E),$$

because $E[2]/2E[2] = (\mathbb{Z}/2\mathbb{Z})^2$. This gives us an upper bound for the rank. As a concrete example of this phenomenon, take

$$E : y^2 = x(x-2)(x-10).$$

Then the set of bad primes is $S = \{2, 5\}$, since it's the primes dividng $\{10, 8, 2\}$. So we get that

$$\operatorname{rank} E(\mathbb{Q}) \leq 2 \cdot 2 = 4.$$

And in fact, in this example the rank is 1.

Remark 6.1.11. The Mordell-Weil theorem is ineffective, in the following sense; we get that $E(\mathbb{Q})$ is finitely generated, but we don't get an explicit algorithm to compute its rank. Why is this? With the example above, we have the injection

$$\phi : E(\mathbb{Q})/2E(\mathbb{Q}) \hookrightarrow \operatorname{Hom}(G_{L/\mathbb{Q}}, (\mathbb{Z}/2\mathbb{Z})^2),$$

where $L = \mathbb{Q}(\sqrt{-1}, \sqrt{2}, \sqrt{p} \ \forall p \mid \Delta_E)$. The Hom group on the RHS is explicit and effective. *The problem:* given ξ in the Hom group, is there a point $P \in E(\mathbb{Q})$ with $\phi(P) = \xi$? We don't have a guaranteed algorithm to answer that question, oddly enough.

Proof of theorem from algebraic number theory. We will first do a few reduction steps. First, note that it's sufficient to take a finite extension K'/K and replace L by $L' = K'L$; similarly, it's okay to make S larger, because in this case we're simply allowing more primes to be ramified. So without loss of generality, we may assume:

1. $\mu_m \subseteq K$.

   Proof of validitiy of assumption 1. Just add in the missing roots of unity. □

2. Making S larger, we may assume that the ring of S-integers R_S is a PID. Recall that this set is defined to be

   $$R_S = \{\alpha \in K : \operatorname{ord}_{\mathfrak{p}}(\alpha) \geq 0 \ \forall \mathfrak{p} \notin S\}.$$

   For example, $S = \emptyset$ means R_S is the ring of integers in K; it's where there are no primes in the denominator. More generally, R_S is where we localize away from the set of primes in S. Why can we assume R_S is a PID? This depends on the fact that ideal class groups in number fields are finite. More precisely:

   Proof of validitiy of assumption 2. Let $\mathfrak{a}_1, \dots, \mathfrak{a}_h$ be ideals in R representing the distinct ideal classes. (The fact that there are only finitely many ideal classes is a very deep theorem.) Consider the set of prime ideals which divides at least one of the $\mathfrak{a}_i$, $S' = \{\mathfrak{p} \mid \mathfrak{a}_1 \cdots \mathfrak{a}_h\}$. Then in $R_{S'}$, the ideals $\alpha_i R_{S'}$ are principal, hence $R_{S'}$ has no non-principal ideals, because we've killed all the non-trivial ideal classes. So we enlarge S by adding the primes in S'. The point: the fact that $\mathfrak{p} \mid \mathfrak{a}_i$ means that we can take an element in $\mathfrak{a}_1 R_{S'}$; this element is now a unit in R'_S. Since this ideal has a unit in it, it's a principal ideal domain. □

Let's recapitulate. We have a number field $K/\mathbb{Q}$ and we're considering the set of K-rational points $E(K)$. For $m \geq 2$, we constructed an injection

$$E(K)/mE(K) \hookrightarrow \mathrm{Hom}(G(K_{E,m}/K), E[m]),$$

where $K_{E,m} := K([m]^{-1}E(K))$ satisfies:

(a) its Galois group is abelian;

(b) its Galois group has exponent m; and

(c) it is unramified outside the finite set of primes $S = \{\mathfrak{p} \mid m\} \cup \{E \text{ has bad reduction at } \mathfrak{p}\}$.

Our goal is to show that $K_{E,m}/K$ is finite, as this will imply the Weak Mordell-Weil theorem. We'll do this by finishing the proof of Theorem 6.1.9. Last time, without loss of generality, we assumed K contains the m'th roots of unity, and we also assumed that the ring of S-integers R_S is a principal ideal domain. Let us re-explain why we can make this latter assumption. Recall that the ideal class group of R_K is finite, so we can take ideals $\mathfrak{a}_1, \ldots, \mathfrak{a}_h \subseteq R_K$, one for each ideal class. Choose $0 \neq \alpha_i \in \mathfrak{a}_i$. We make S bigger by considering the finite set of primes

$$S' = \bigcup_{i=1}^{h} \{\mathfrak{p} \subseteq K : \mathrm{ord}_{\mathfrak{p}}(\alpha_i) \neq 0\}.$$

And notice that in the ring of S' integers, $R_{S'}$, the ideal classes $\mathfrak{a}_i$ disappear; concretely, we can see that

$$\mathfrak{a}_i R_{S'} = R_{S'}.$$

This follows because $\alpha_i \in \mathfrak{a}_i$ and $\alpha_i \in R_{S'}^*$ (why is α_i a unit in $R_{S'}$? It's not divisible by any of the prime ideals, i.e. it has valuation zero for all relevant primes.)

Abstractly: consider the maximum abelian extension K'/K with exponent m. Because $\mu_m \subseteq K$, if $a \in K^*$, then $k : G(K(a^{1/m})/K) \hookrightarrow \mathbb{Z}/m\mathbb{Z}$ is an injection defined by $\sigma(a^{1/m}) = \zeta_m^{k(\sigma)} a^{1/m}$. But the main theorem of Kummer theory says that this is the only way to generate abelian extensions; more concretely, Kummer theory says that the maximal extension is exactly

$$K' = K(\sqrt[m]{a} : a \in K^*).$$

But if we replace a with ab^m, where $b \in K$, then we get the same field; so the right way to define this is

$$K' = K(\sqrt[m]{a} : a \in K^*/(K^*)^m).$$

At what primes is $K(a^{1/m})/K$ ramified? This is the splitting field of $X^m - a = 0$. When will two roots of this come together modulo $\mathfrak{p}$? This happens if $\sqrt[m]{a} - \zeta \sqrt[m]{a} \equiv 0 \pmod{\mathfrak{p}}$, where $1 \neq \zeta \in \mu_m$. There are two possibilities; $\mathfrak{p} \mid \mathfrak{a}$, or $\mathfrak{p} \mid 1 - \zeta$ (but this latter condition implies $\mathfrak{p} \mid m$.) Alternative computation of those values of $\mathfrak{p}$: the discriminant of the splitting polynomial is

$$\mathrm{Disc}(X^m - a) = \mathrm{Res}(X^m - a, mx^{m-1}) = \mathrm{Res}(X^m - a, m) \cdot \mathrm{Res}(X^m - a, X^{m-1}) = m^m \cdot (-a)^{m-1},$$

so the only potential ramification is primes dividing m and primes dividing a. In summary, $K(\sqrt[m]{a})/K$ is ramified at most at the primes:

1. $\mathfrak{p} \mid m$
2. $\mathfrak{p} \mid a$; but $\mathfrak{p} \mid a$ and $\mathfrak{p} \nmid m$ if and only if $\mathrm{ord}_{\mathfrak{p}}(a) \neq 0 \pmod{m}$ (because if a is an m'th power and we take an m'th root, we're not getting any ramification.)

The net result is that $L \subseteq K(\sqrt[m]{a} : a \in K^*/(K^*)^m)$, but L is unramified outside of S. So let us consider

$$T_S := \{a \in K^*/(K^*)^m : \mathrm{ord}_{\mathfrak{p}}(a) \equiv 0 \pmod{m} \forall \mathfrak{p} \notin S\}.$$

Then the field L will be obtained by taking m'th roots of elements in T_S, i.e.

$$L = K(\sqrt[m]{a} : a \in T_S).$$

We claim that T_S is finite; once we show this, we're done.

How can we prove that T_S is finite? Recall that $R_S^* = \{a \in R_K : \mathrm{ord}_{\mathfrak{p}}(a) = 0 (\forall \mathfrak{p} \notin S)\}$. Certainly there is a map

$$R_S^* \to T_S,$$

because valuation 0 implies valuation is congruent to zero. We claim this map is surjective. To see this, let $a \in T_S$. Therefore, the principal ideal $aR_S = \mathfrak{b}^m$ for some ideal $\mathfrak{b} \in R_S$; here, we're using the fact that in a Dedekind domain, every ideal has a unique factorization into prime ideals. But R_S is a principal ideal domain by construction; so this implies that $\mathfrak{b}$ is a principal ideal, so $\mathfrak{b} = bR_S$ for some b. This implies that $aR_S = b^m R_S$, so $a = ub^m$ for some unit $u \in R_S^*$. This implies that $a = u$ in $T_S \subseteq K^*/(K^*)^m$. And of course $u \in R_S^*$. Surjectivity of the map $R_S^* \to T_S$ follows. We should point out that $(R_S^*)^m$ is in the kernel of this map, i.e., m'th powers get mapped to 1. So what we actually get is the surjection

$$R_S^*/(R_S^*)^m \to T_S.$$

Now, recall Dirichlet's unit theorem, which says that R_K^* is a finitely generated abelian group, and $R_K^* \cong \mu(K) \times \mathbb{Z}^{r_1+r_2-1}$, where $\mu(K)$ is the roots of unity in K, r_1 is the number of real embeddings, and r_2 is half the number of complex embeddings. In fact, an extension of Dirichlet's unit theorem says that

$$R_S^* \cong \mu(K) \times \mathbb{Z}^{r_1+r_2+\#S-1}.$$

This immediately implies that $R_S^*/(R_S^*)^m$ is finite, as $\mathbb{Z}/m\mathbb{Z}$ is finite, and in fact

$$\#(R_S^*/(R_S^*)^m) \approx m^{r_1+r_2+\#S},$$

as needed. □

This completes the proof of the weak Mordell-Weil theorem, which relied crucially on two finiteness results from algebraic number theory: finiteness of the class group, and the finite generation of the group of units. □

Remark 6.1.12. Say E is defined over K, and $K_m = K(E[m])$. Then we can trace this argument through and get an upper bound

$$\mathrm{rank}\, E(K) \leq \text{some function of } [K_m : K] \text{ and the } m\text{-torsion in the ideal class group of } R_{K(E[m])}.$$

6.2 Mordell-Weil theorem

First, some motivation for the proof. We now know that $E(K)/mE(K)$ is finite. And we want to use this to prove that $E(K)$ is finitely generated. We'll need an extra piece of information for this deduction; for example, $A = \mathbb{Q}$ is an abelian group, and A/mA is finite for all $m \geq 1$ (in fact it's 0) but A is not a finitely generated abelian group. If you try to use finiteness of the quotient groups to get finite generation of the larger group, then we get that there are $a \in A$ which are infinitely m-divisible; i.e. for every m^k there exists $b \in A$ so that $m^k b = a$. In order to get around this problem in the elliptic curves case, we want to know that elements of $E(K)$ cannot be divided by arbitrarily high powers of $[m]$. So the rough idea: given $P \in E(K)$, if $P = mQ$ for some $Q \in E(K)$, then somehow Q is "less complicated" than Q; for example, in computer science language, you can describe it using fewer bits. An analogy: take $0 \neq a \in \mathbb{Z} \subseteq \mathbb{Q}^*$. If $a = b^m$ for some $b \in \mathbb{Z}$, then b is less complicated than a, because

$$\#\text{bits in } b \approx \frac{\#\text{ bits in } a}{m}.$$

Can we repeat this process indefinitely, while remaining inside $\mathbb{Z}$? The answer is no (unless we started with $a = 1$) because the number of bits is necessarily a whole number. This strategy is related to Fermat's notion of infinite descent. The theorem that we'll prove:

Theorem 6.2.1. *Let A be an abelian group, let $m \geq 2$, and suppose we have a way of measuring the complexity of elements of A, i.e. suppose we have a function $h : A \to [0, \infty)$ satisfying:*

1. *$h(ma) \geq m^2 h(a) - C_1(A, m)$. This says "multiplying an element by m makes it a lot more complicated."*
2. *$h(a + b) \leq 2h(a) + C_1(A, b)$.*
3. *$\{a \in A : h(a) \leq B\}$ is finite. This says there are only finitely many things of bounded complexity, i.e. "you can't continually make things less and less complex."*

Then A is a finitely generated abelian group.

(Lecture 28: March 26, 2021)

Now we re-state that result in a more precise way.

Proposition 6.2.2 (Descent proposition)**.** *Let A be an abelian group, let $m \geq 2$, and suppose there exists a functon $h : A \to [0, \infty]$ such that:*

(a) A/mA is finite.

(b) $h(mP) \geq m^2 h(P) - C_1$ for all $P \in A$.

(c) $h(P + Q) \leq 2h(P) + C_2(Q)$ for all $P \in A$.

(d) $\{P \in A : h(P) \leq X\}$ is finite for all X.

Then A is finitely generated.

In number theoretic applications, we generally take h to be the *height.*

Proof of descent proposition. Choose coset representatives $Q_1, \ldots, Q_r$ for A/mA. Let $P \in A$ be arbitrary. The goal is to write P as a $\mathbb{Z}$-linear combination of $Q_1, \ldots, Q_r$ plus an element of a set of bounded height. Write $P_0 := P$. We can write $P_0 = mP_1 + Q_{i_1}$ by looking at the coset of P in A/mA. Then we can write $P_1 = mP_2 + Q_{i_2}$, and inductively, until we get $P_{n-1} = mP_n + Q_{i_n}$. For any j, we have $P_{j-1} = mP_j + Q_{i_j}$. How big is P_j compared to P_{j-1}? We have

$$\begin{aligned} h(P_j) &\leq_{(b)} \frac{1}{m^2} \left(h(mP_j) + C_1\right) \\ &= \frac{1}{m^2} \left(h(P_{j-1} - Q_{i_j}) + C_1\right) \\ &\leq_{(c)} \frac{1}{m^2} (2h(P_{j-1}) + C_2(-Q_{i_j}) + C_1) \\ &\leq \frac{2}{m^2} h(P_{j-1}) + C, \end{aligned}$$

where $C := C_1 + \max_{1 \leq i \leq r}\{C_2(-Q_i)\}$. What this says is that the height of P_j is significantly less than the

height of P_{j-1}, since $m \geq 2$. Inductively, we can use the above to estimate that

$$\begin{aligned} h(P_n) &\leq \frac{2}{m^2} h(P_{n-1}) + C \\ &\leq \frac{2}{m^2}\left(\frac{2}{m^2} h(P_{n-2}) + C\right) + C \\ &= \left(\frac{2}{m^2}\right)^2 h(P_{n-2}) + \left(1 + \frac{2}{m^2}\right) C \\ &\leq \left(\frac{2}{m^2}\right)^2 \left(\frac{2}{m^2} h(P_{n-3}) + C\right) + \left(1 + \frac{2}{m^2}\right) C \\ &= \left(\frac{2}{m^2}\right)^3 h(P_{n-3}) + \left(1 + \frac{2}{m^2} + \left(\frac{2}{m^2}\right)^2\right) C, \end{aligned}$$

and eventually,

$$h(P_n) \leq \left(\frac{2}{m^2}\right)^n h(P_0) + \frac{1}{1 - \frac{2}{m^2}} C \leq \left(\frac{2}{m^2}\right)^n h(P) + 2C \leq \frac{1}{2^n} h(P) + 2C$$

because $m \geq 2$. Why is this good? We started with an arbitrary point P, and we repeatedly divided by m until we obtained a point whose height we can get arbitrarily small by taking n large enough. If we take $n = \lceil \log_2(h(P) + 1) \rceil$, then the above estimate implies

$$h(P_n) \leq 1 + 2C.$$

On the other hand, if we do a bunch of back-substituting, then we get

$$P = mP_1 + Q_{i_1} = m(mP_2 + Q_{i_2}) + Q_{i_1} = \cdots = m^n P_n + \sum_{j=1}^{n} m^{j-1} Q_{i_j}.$$

This shows that $P \in \mathrm{Span}_{\mathbb{Z}}\{P_n, Q_1, \ldots, Q_r\}$. But we know that P_n has bounded height, so in fact

$$\mathrm{Span}_{\mathbb{Z}}\{P_n, Q_1, \ldots, Q_r\} \subseteq \mathrm{Span}_{\mathbb{Z}}\{Q_1, \ldots, Q_r\} \cup \{R \in A : h(R) \leq 2C + 1\}.$$

Note that $\{R \in A : h(R) \leq 2C+1\}$ is finite, and independent of P, so the above is indeed a finite generating set for A. □

Remark 6.2.3. The reason the Mordell-Weil theorem is not effective is that we don't actually know how to find coset representatives of $E(K)/mE(K)$, i.e. we don't have an algorithm to find $Q_1, \ldots, Q_r$.

Thanks to this result, in order to prove Mordell-Weil, we need to construct a height function on $E(K)$.

6.3 Height functions

Motivation: how do we measure the complexity of numbers? For $a/b \in \mathbb{Q}$, a good way to measure is $\max\{|a|, |b|\}$. But notice that there is ambiguity here; this says $1/2$ is less complicated than $50/100$. So we must insist that $\gcd(a, b) = 1$, i.e. that a/b be in lowest terms. More generally, consider $P = [a_0, \ldots, a_N] \in \mathbb{P}^N(\mathbb{Q})$. If we normalize so that $a_i \in \mathbb{Z}$ and $\gcd(a_i) = 1$, then we can define the height of P to be $H(P) = \max |a_i|$.

Claim 6.3.1. $\#\{P \in \mathbb{P}^n(\mathbb{Q}) : H(P) \leq X\}$ is finite.[12]

Proof. Each coordinate has at most $2X + 1$ possibilites, and there are $N + 1$ coordinates, so an upper bound is given by $(2X + 1)^{N+1}$. □

[12] In fact, one can show that $\#\{P \in \mathbb{P}^N(\mathbb{Q}) : H(P) \leq X\} \sim cX^N$ if $N \geq 2$.

If you were to try this same strategy with number fields, then you would replace $a_i \in \mathbb{Z}$ with $a_i \in R_K$, which is a problem in general, because R_K is not necessarily a unique factorization domain. One way of getting around this is to work with fractional ideals and taking norms. This is easy to set up but harder to prove things about. Instead, we'll follow Weil's strategy. The idea: $\mathbb{Q}$ has the absolute value $|\cdot|$ when we view $\mathbb{Q} \subseteq \mathbb{R}$, as well as the absolute value $|\cdot|_p$ when we view $\mathbb{Q} \subseteq \mathbb{Q}_p$.

Definition 6.3.2. Let $M_{\mathbb{Q}}$ be the set of *standard absolute values* on $\mathbb{Q}$, so it has:

1. The archimedian absolute value $|x|_\infty = \max\{x, -x\}$.
2. For each prime p, the p-adic absolute value defined by $|x|_p = p^{-n}$; here, $x = p^n \cdot \frac{a}{b}$ where $n \in \mathbb{Z}, a, b \in \mathbb{Z}, b \neq 0, p \nmid a, p \nmid b$. We also define $|0|_p = 0$.

Note: for any $x \in \mathbb{Q}$, we have the *product formula*

$$\prod_{v \in M_{\mathbb{Q}}} |x|_v = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0. \end{cases}$$

Next, we'll find the standard absolute values on a number field.

(Lecture 29: March 29, 2021)

Definition 6.3.3. For any number field $K/\mathbb{Q}$, we define

$$M_K := \{\text{absolute values on } K \text{ extending the ones in } M_{\mathbb{Q}}\}.$$

Recall that if an absolute value satisifes the ultrametric inequality $|x + y| \leq \max\{|x|, |y|\}$, then we say it's *nonarchimedian*. We write:

1. M_K^∞ denotes the set of archimedian absolute values; we get one for each real embedding $K \hookrightarrow \mathbb{R}$, and we get one for each pair of complex embeddings $K \hookrightarrow \mathbb{C}$.
2. M_K^0 denotes the set of nonarchimedian absolute values; for each $\mathfrak{p}R_K$, we get a $\mathfrak{p}$-adic absolute value $|x|_{\mathfrak{p}}$ which extends the p-adic absolue value of primes $\mathfrak{p}$ lying over p. In other words, lying above p, there are many absolute values, corresponding to all the prime factors in the splitting $p = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$ in R_K. Ignoring a normalization factor, such absolute values are defined as

$$|x|_{\mathfrak{p}} = (N\mathfrak{p})^{-t}, \qquad \text{where } xR_K = \mathfrak{p}^t\sigma.$$

For $v \in M_K$, we write the local degree as

$$n_v := [K_v : \mathbb{Q}_v].$$

Note that $\mathbb{Q}_v$ will be either $\mathbb{R}$ of $\mathbb{Q}_p$, and in particular, this is a finite extension; furthermore, we have that $n_v \in \{1, \ldots, [K : \mathbb{Q}]\}$. Some basic formulas from algebraic number theory:

1. Given $L/K/\mathbb{Q}$ with $v \in M_K$, we have

$$\sum_{\substack{w \in M_L \\ v|w}} n_w = [L : K] n_v.$$

2. Given $x \in K^*$, we have

$$\prod_{v \in M_K} |x|_v^{n_v} = 1.$$

Some books define $\|x\|_v := |x|_v^{n_v}$.

Definition 6.3.4. Let $K/\mathbb{Q}$ be a number field, and consider a point $P = [x_0, \ldots, x_N] \in \mathbb{P}^N(K)$. The *height of P (relative to K)* is

$$H_K(P) := \prod_{v \in M_K} \max\{|x_0|_v, \ldots, |x_N|_v\}^{n_v}.$$

Various facts about relative height:

(a) $H_K(P)$ is independent of homogeneous coordinates for K.

Proof. If $P = [\alpha x_0, \dots, \alpha x_N]$ for some $\alpha \in K^*$, then

$$\prod_{v \in M_K} \max\{|\alpha x_0|_v, \dots, |\alpha x_N|_v\}^{n_v} = \prod_{v \in M_K} |\alpha|_v^{n_v} \max\{|x_0|_v, \dots, |x_N|_v\}^{n_v} = \prod_{v \in M_K} \max\{|x_0|_v, \dots, |x_N|_v\}^{n_v}$$

by the product formula. □

(b) $H_K(P) \geq 1$.

Proof. $P = [x_0, \dots, x_N] \in \mathbb{P}^N(K)$, then some $x_j \neq 0$, so $P = [x_0/x_j, \dots, 1, \dots, x_N/x_j]$. This implies that

$$H_K(P) = \prod_{v \in M_K} \max\{|x_0/x_j|_v, \dots, |1|_v, \dots, |x_N/x_j|\}^{n_v} \geq 1.$$

□

(c) Given an extension L/K and a point $P \in \mathbb{P}^N(K)$, then

$$H_L(P) = H_K(P)^{[L:K]}.$$

The proof of this just uses the formula explaining how the local degrees change. A consequence of this: if you take an appropriate root, then you can get a height function that doesn't depend on the field.

Definition 6.3.5. The (absolute) *Weil height* on $\mathbb{P}^n(\overline{\mathbb{Q}})$ is a function

$$H : \mathbb{P}^N(\overline{\mathbb{Q}}) \to [1, \infty)$$

which is defined as follows: for any $P \in \mathbb{P}^N(\overline{\mathbb{Q}})$, choose any number field $K/\mathbb{Q}$ with $P \in \mathbb{P}^N(K)$. Then

$$H(P) := H_K(P)^{1/[K:\mathbb{Q}]}.$$

Recall that a morphism is a map $F : \mathbb{P}^N \to \mathbb{P}^M$ given by $F(P) = [f_0(P), \dots, f_M(P)]$, where $f_0, \dots, f_M \in \overline{\mathbb{Q}}[x_0, \dots, x_N]$ are homogeneous polynomials with no common zeros in $\mathbb{P}^N$. The degree of the rational map is $\deg F = \deg f_i$. And we say F is defined over K if $f_i \in K[x_0, \dots, x_n]$. Our next goal is to show that heights transform nicely relative to morphisms.

Proposition 6.3.6. *Let $F : \mathbb{P}^N \to \mathbb{P}^M$ be a morphism of degree $d \geq 1$. Then there are constants $C_1, C_2 > 0$ (depending on F, N, M) such that for all $P \in \mathbb{P}^N(\overline{\mathbb{Q}})$,*

$$C_1 H(P)^d \leq H(F(P)) \leq C_2 H(P)^d.$$

Note that the upper bound holds for all rational maps, but the lower bound is often false for rational maps which are not morphisms; in fact, measuring how false it is is a very interesting question. A fun exercise: show that the lower bound is false for $F([x, y, z]) = [x^2, xy, z^2]$.

Proof of proposition. Consider the map $F = [f_0, \dots, f_M]$ and the point $P = [x_0, \dots, x_N] \in \mathbb{P}^N(\overline{\mathbb{Q}})$. Fix $K/\mathbb{Q}$ with $P \in \mathbb{P}^N(K)$ and $f_0, \dots, f_m \in K[x_0, \dots, x_N]$. Some notation:

- We'll write $|P|_v := \max\{|x_0|_v, \dots, |x_N|_v\}$. So we have $|F(P)|_v = \max\{|f_0(P)|_v, \dots, |f_N(P)|_v\}$.
- It'll be convenient to write $|F|_v := \max\{|a|_v : a \text{ is a coefficient of one of the } f_i\}$. This is convenient because $H_K(P) = \prod_{v \in M_K} |P|_v^{n_v}$, and similarly $H_K(F(P)) = \prod_{v \in M_K} |F(P)|_v^{n_v}$.
- Given these notations, it makes sense to define $H(F) := \prod_{v \in M_K} |F|_v^{n_v}$. Note that this is an effective constant; given an explicit map, we can actually compute this quantity.

- We will use constants $c_1, c_2, \dots$ that depend on $F, N, M, d = \deg F$.
- We will define
$$\epsilon(v) := \begin{cases} 1 & v \in M_K^\infty \\ 0 & v \in M_K^0. \end{cases}$$
We can use this to write a more general version of the triangle inequality:
$$|t_1 + \cdots + t_n|_v \le n^{\epsilon_v} \max\{|t_1|_v, \dots, |t_n|_v\}.$$
If $v \in M_K^\infty$, then this is slightly weaker than the triangle inequality; and if $v \in M_K^0$ then this is exactly the non-archimedian absolute value.

Now to the proof. We can estimate
$$\begin{aligned} |f_i(P)|_v &= \left| \sum_{\substack{e=(e_0,\dots,e_N) \\ e_0+\cdots+e_N=d}} a_{i,e} x_0^{e_0} \cdots x_n^{e_N} \right|_v \\ &\le (\#\text{terms in the sum})^{\epsilon(v)} \cdot (\max_e\{|a_{i,e}|_v\}) \cdot (\max_e\{|x_0^{e_0} \cdots x_N^{e_N}|\}) \\ &\le \binom{N+d}{d}^{\epsilon(v)} |F|_v |P|_v^d. \end{aligned}$$

If we raise both sides to the n_v power, and then take the maximum over $0 \le i \le M$, and multiply over all $v \in M_K$, and take the $1/[K:\mathbb{Q}]$ root (so that we get absolute heights instead of heights relative to K) then we obtain that
$$\begin{aligned} H(F(P)) &\le \left(\prod_{v \in M_K} c_1^{\epsilon(v) n_v/[K:\mathbb{Q}]} \right) \left(\prod_{v \in M_K} |F|_r^{n_v} \right)^{1/[K:\mathbb{Q}]} \left(\prod_{v \in M_K} |P|_v^{d n_v} \right)^{1/[K:\mathbb{Q}]} \\ &= c_2 H(F) H(P)^d \\ &= c_1^{\left(\sum_{v \in M_K^\infty} n_v\right)/[K:\mathbb{Q}]} H(F) H(P)^d, \end{aligned}$$
but that quotient in the exponent is 1, so this implies
$$H(F(P)) \le c_1(N,d) H(F) H(P)^d.$$

(Lecture 30: March 31, 2021)

Today we'll prove the lower bound. Recall that $F = [f_0, \dots, f_N]$ is a morphism, which implies that
$$\{Q \in \overline{\mathbb{Q}}^{N+1} : f_0(Q) = \cdots = f_M(Q) = 0\} = \{(0, \dots, 0)\} = \{Q \in \mathbb{Q}^{N+1} : x_0(Q) = \cdots = x_N(Q) = 0\}$$
By the Nullstellensatz, this implies that $\sqrt{\langle f_0, \dots, f_M \rangle} = \sqrt{\langle x_0, \dots, x_N \rangle}$ in $\overline{\mathbb{Q}}[x_0, \dots, x_N]$, where $\sqrt{I} = \{\alpha \in I : \alpha^n \in I \text{ for some } n \ge 1\}$. Hence for each i, there exists e so that $x_i^e \in \langle f_0, \dots, f_M \rangle$. Let e be the largest e's for the various i's. This tells us that
$$x_i^e = \sum_{j=1}^M g_{ij} f_j \text{ for some } g_{ij} \in K[x_0, \dots, x_N].$$
The f_j are homogeneous of degree d, so x_i^e is homogeneous of degree e, so without loss of generality, we may assume the g_{ij} are homogeneous of degree $e - d$. Denote by $|G|_v$ the maximum of the $|\cdot|_v$-valuations of the coefficients of the g_{ij}. Let $H_K(G) = \prod_{v \in M_K} |G|_v^{n_v}$. The point here is that $|G|_v$ and $H_K(G)$ are independent of P, but depend only on F. The triangle inequality applied to the above implies that
$$|x_i(P)|_v^e \le M^{\epsilon(v)} \max_{i,j}\{|g_{ij}(P)|_v |f_j(P)|_v\} \le M^{\epsilon(v)} \max_{i,j}\{|g_{ij}(P)|\} \max_j\{|f_j(P)|_v\},$$

where $\epsilon(v) = 1$ if v is archimedian, and $\epsilon(v) = 0$ if v is non-archimedian. But we can estimate that

$$\begin{aligned}|g_{ij}(P)|_v &\le (\#\text{of terms in } g_{ij})^{\epsilon(v)} \cdot \max|\text{coeffs of } g_{ij}|_v \cdot \max|\text{coeffs of } P|_v^{\deg g_{ij}} \\ &\le c_3^{\epsilon(v)}|G|_v|P|_v^{e-d},\end{aligned}$$

So we can continue to estimate that

$$|P|_v^e = \max_i |x_i(P)|_v^e \le c_3^{\epsilon(v)}|G|_v|P|_v^{e-d}|F(P)|_v \implies |P|_v^d \le c_3^{\epsilon(v)}|G|_v|F(P)|_v.$$

Raising both sides to the n_v power, taking the product over $v \in M_K$ so as to get heights, then taking the $1/[K:\mathbb{Q}]$ root to get the absolute height, we get that

$$H(P)^d \le c_3 H(G)H(F(P)).$$

This completes the proof. The takeaway is that the Nullstellensatz allowed us to flip the inequality. □

That tells us that heights behave in a nice way when we apply morphisms. Our next task is to show that heights actually measure a sort of complexity. Some notation: for $x \in \overline{\mathbb{Q}}$, we write $H(x) := H([x,1])$.

Proposition 6.3.7. *If we write $f(T) \in \overline{\mathbb{Q}}(T)$ as*

$$f(T) = a_0T^d + \cdots + a_d = a_0(x-\alpha_1)\cdots(x-\alpha_d),$$

then

$$2^{-d}\prod_{j=1}^d H(\alpha_i) \le H([a_0,\ldots,a_d]) \le 2^{d-1}\prod_{j=1}^d H(\alpha_i).$$

This says that the coefficients have big height if and only if the roots have big height. The proof of this is in the textbook. Next we'll show that the height is Galois invariant.

Proposition 6.3.8. *Let $P \in \mathbb{P}^n(\overline{\mathbb{Q}})$ and $\sigma \in G_{\overline{\mathbb{Q}}/\mathbb{Q}}$. Then $H(P^\sigma) = H(P)$.*

Proof. Fix a Galois extension $K/\mathbb{Q}$ so that $P \in \mathbb{P}^N(K)$, and consider $\sigma \in G_{K/\mathbb{Q}}$. Observe that σ induces a bijection on M_K, defined via $\sigma(v)$ is $|\alpha|_{\sigma(v)} = |\alpha^\sigma|_v$. Then $n_{\sigma(v)} = [K_{\sigma(v)} : \mathbb{Q}_p]$, and $K_{\sigma(v)} \cong K_v$ via the isomorphism $\alpha \mapsto \sigma^{-1}(\alpha)$. So we can compute that

$$H_K(P^\sigma) = \prod_{v\in M_K}\max|x_i^\sigma|_v^{n_v} = \prod_{v\in M_K}\max|x_i|_{\sigma(v)}^{n_{\sigma(v)}} = \prod_{w\in M_K}\max|x_i|_w^{n_w} = H_K(P),$$

where we used the change of variables $w \leftrightarrow \sigma(v)$. This completes the proof. □

We now have all the tools to prove Northcott's theorem:

Theorem 6.3.9 (Northcott)**.** *Fix constants C, D. Then*

$$\{P \in \mathbb{P}^N(\overline{\mathbb{Q}}) : H(P) \le C, [\mathbb{Q}(P):\mathbb{Q}] \le D\}$$

is a finite set.

Proof. Let $P = [x_0,\ldots,x_N] \in \mathbb{P}^N(\overline{\mathbb{Q}})$. There exists j with $x_j \neq 0$, so without loss of generality assume $x_j = 1$. Then we can estimate that

$$H_{\mathbb{Q}(P)}(P) = \prod_{v\in M_{\mathbb{Q}(P)}}\max_i|x_i|_v^{n_v} \ge \max_{0\le i\le N}\prod_{v\in M_{\mathbb{Q}(P)}}\max\{|x_i|_v^{n_v},1\} = \max_i H_{\mathbb{Q}(P)}(x_i).$$

In summary, we've shown that $H(P) \le C$ implies $H(x_i) \le C$ for all $0 \le i \le N$, and of course $[\mathbb{Q}(x_i):\mathbb{Q}] \le [\mathbb{Q}(P):\mathbb{Q}]$. Therefore it suffices to show that the following set is finite:

$$\{x \in \overline{\mathbb{Q}} : H(x) \le C, [\mathbb{Q}(x):\mathbb{Q}] = D\}.$$

Let α be in this set, and let $\alpha_1, \dots, \alpha_d$ be the $G_{\overline{\mathbb{Q}}/\mathbb{Q}}$ conjugates of α. Then the minimal polynomial of α is

$$f_\alpha = \prod_{j=1}^{d} (x - \alpha_j) = T^D + a_1 T^{D-1} + \cdots + a_d.$$

If we write $H(f) = H([1, a_1, \dots, a_D])$, then we can estimate using the previous propositions that

$$H(f) \leq 2^{D-1} \prod_{j=1}^{D} H(\alpha_j) \leq 2^{D-1} H(\alpha)^D \leq 2^{D-1} C^D.$$

So we've proven that

$$\{\alpha : H(\alpha) \leq C, [\mathbb{Q}(\alpha) : \mathbb{Q}] = D\} \to \{a \in \mathbb{P}^D(\mathbb{Q}) : H(a) \leq 2^{D-1} C^D\}$$

is at most a D-to-1 map; but the latter set is finite, which finishes the proof. □

(Lecture 31: April 2, 2021)

Northcott's theorem can be used to prove a very important theorem of Kronecker. Because

$$H(\alpha) = \Big(\prod_{v \in M_{\mathbb{Q}(\alpha)}} \max\{\|\alpha\|_v, 1\}^{n_v} \Big)^{1/[\mathbb{Q}(\alpha):\mathbb{Q}]},$$

it's clear that $H(\alpha) \geq 1$. We have a converse theorem as well:

Theorem 6.3.10 (Kronecker)**.** *Let $0 \neq \alpha \in \overline{\mathbb{Q}}$. Then $H(\alpha) = 1$ if and only if α is a root of unity.*

Proof. If $\alpha^n = 1$, then $\alpha_v^n = 1$, so $\alpha_v = 1$ for all v, so $H(\alpha) = 1$. Conversely, because $H(\alpha^n) = H(\alpha)^n$, we have that $H(\alpha) = 1$ implies $H(\alpha^n) = 1$ for all n, which implies that $\{\alpha^n : n \geq 0\}$ is a set of bounded height and lives in $\mathbb{Q}(\alpha)$, which is an extension of $\mathbb{Q}$ of finite degree, so by Northcott, we have there exists $i > j$ so that $\alpha^i = \alpha^j$, hence $\alpha^{i-j} = 1$. □

This theorem brings up a natural question: if $\alpha \neq 0$ is not in μ, then $H(\alpha) > 1$ by Kronecker, so it is natural to ask how close $H(\alpha)$ can get to 1. Well, we can get fairly close, because

$$H(2^{1/n}) = H(2)^{1/n} = 2^{1/n} \to 1$$

as $n \to \infty$. But by an observation of Lehmer, in order to get numbers with height close to 1, we need to take numbers in bigger and bigger number fields. There is a conjecture that you can't do better than this trick.

Conjecture 6.3.11 (Lehmer)**.** *There is an absolute constant $C > 0$ so that for all $\alpha \in \overline{\mathbb{Q}}$ which are nonzero and are not roots of unity, we have that*

$$\log H(\alpha) \geq \frac{C}{[\mathbb{Q}(\alpha) : \mathbb{Q}]}.$$

The best known result is due to Dobrovolwki, in the 1970's:

Theorem 6.3.12 (Dobrovolwki)**.** *We have*[13]

$$\log H(\alpha) \geq \frac{C}{[\mathbb{Q}(\alpha) : \mathbb{Q}]} \left(\frac{\log \log [\mathbb{Q}(\alpha) : \mathbb{Q}]}{\log [\mathbb{Q}(\alpha) : \mathbb{Q}]} \right)^3.$$

[13] Joe says: "I will give you an A in the course, as well as a PhD, if you can decrease this exponent from 3 to 2."

6.4 Completing the proof of the Mordell-Weil theorem

First, some notation. Given functions $f, g : S \to \mathbb{R}$, we write $f = g + O(1)$ if there exists some $C = C_{f,g,S}$ such that $|f(x) - g(x)| \leq C$ for all $x \in S$. Also, we define

$$h(P) := \log H(P),$$

so in particular, $h(P) \geq 0$, and $h(\alpha) = 0$ if and only if $\alpha = 0$ or $\alpha \in \mu$.

To finish the proof of Mordell-Weil, we need a height function $h : E(\overline{K}) \to \mathbb{R}$. We will simply use the x-coordinate. Fix a Weierstrass equation $E : y^2 = x^3 + Ax + B$, and we'll define the height of a point P to be

$$h(P) := h_E(P) := \begin{cases} h(x(P)) & P \neq \mathcal{O} \\ 0 & P = \mathcal{O}. \end{cases}$$

What we're really doing is taking $x : E \to \mathbb{P}^1$ and taking $h_E(P) = h(x(P))$.

Proposition 6.4.1. *$\{P \in E(K) : h(P) \leq R\}$ is finite.*

Proof. We can map

$$\{P \in E(K) : h(P) \leq R\} \to \{\alpha \in \mathbb{P}^1(K) : h(\alpha) \leq R\} : P \mapsto x(P).$$

The latter set is finite by Northcott. And this map is at most two-to-one, because if we fix the x-coordinate, there are at most two corresponding y-coordinates. □

We also must show that $h(mP) \geq m^2 h(P) - C$, and $h(P + Q) \leq 2h(P) + C'_Q$. We'll prove something much stronger than that: we'll prove a sort of parallelogram law, and we'll get what we need from that using induction.

Theorem 6.4.2. *For all $P, Q \in E(K)$, we have*

$$h(P + Q) + h(P - Q) = 2h(P) + 2h(Q) + O_E(1).$$

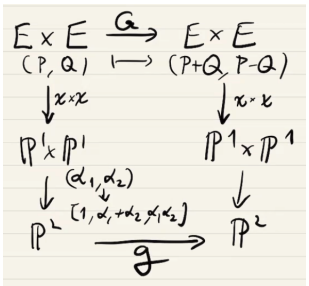
Sketch of proof. Let $x_1 = x(P), x_2 = x(Q), x_3 = x(P + Q), x_4 = x(P - Q)$. Using the doubling formula, one can compute that

$$x_3 + x_4 = \frac{2(x_1 + x_2)(A + x_1x_2) + 4B}{(x_1 + x_2)^2 - 4x_1x_2}, \qquad x_3x_4 = \frac{(x_1x_2 - A)^2 - 4B(x_1 + x_2)}{(x_1 + x_2)^2 - 4x_1x_2}.$$

Define $[t, u, v] = [1, x_1 + x_2, x_1x_2]$. We can map $[1, x_1 + x_2, x_1x_2] \mapsto [1, x_3 + x_4, x_3x_4]$, so there should be a corresponding function of $[t, u, v]$ that reflects this; one can compute that it is

$$g([t, u, v]) = [u^2 - 4tv, 2u(At + v) + 4Bt^2, (v - At)^2 - 4Btu].$$

Then g makes the following diagram commute,



as it tells us how to take symmetric functions of x_1, x_2 and turn them into symmetric functions of x_3, x_4. Say σ is the composition $E \times E \to \mathbb{P}^2$ down the left side.

Now, we want to apply the height transformation formula to g, but in order to do that, we need to show that $g : \mathbb{P}^2 \to \mathbb{P}^2$ is a morphism, which we can do directly. We will assume

$$[u^2 - 4tv, 2u(At + v) + 4Bt^2, (v - At)^2 - 4Btu] = [0, 0, 0]$$

and show that this forces $t = u = v = 0$. If $t = 0$, then the first coordinate implies $u = 0$, so the third coordinate gives $v = 0$. If $t \neq 0$, let $x = u/2t$; we divide through by $4t^2$, so the first coordinate becomes $x^2 = v/t$, the second becomes $4x^3 + 4Ax + 4B = 0$, and the third becomes $x^4 - 2Ax - 8Bx + A^2 = 0$. But we know that $x(2P) = (4x^3 + 4Ax + 4B)/(x^4 - 2Ax - 8Bx + A^2)$, and in fact these polynomials have no common roots; one can check that the resultant of these is $(4A^3 + 27B^2)^2$, which is not zero. So there are no common roots in the second and third coordinates. It follows that g actually is a morphism.

Returning to the commutative diagram, we can compute that

$$h(\sigma(P - Q, P + Q)) = h(\sigma(G(P, Q))) = h(g(\sigma(P, Q))) = 2h(\sigma(P, Q)) + O(1)$$

because g is a morphism on $\mathbb{P}^2$, and $\deg g = 2$. This implies that

$$h([1, \alpha + \beta, \alpha\beta]) = h(\alpha) + h(\beta) + O(1),$$

because α and β are the roots of $T^2 - (\alpha + \beta)T + \alpha\beta$. This implies that $h(x_3) + h(x_4) = 2h(x_1) + 2h(x_2) + O(1)$, but $h(x_3) = h(P + Q)$ and $h(x_4) = h(P + Q)$, and $h(x_1) = h(P)$ and $h(x_2) = x(Q)$. □

(Lecture 32: April 5, 2021)

6.5 Finishing the proof of the Mordell-Weil theorem

Let E/K be an elliptic curve with Weierstrass equation $y^2 = x^3 + Ax + B$. We defined a logarithmic height function

$$h : E(K) \to \mathbb{R} : h(P) = h(x(P)).$$

We know $h(P) \geq 0$, and $\{P \in E(K) : h(P) \leq C\}$ is finite. The main theorem that we proved last time was a sort of parallelogram law,

$$h(P + Q) + h(P - Q) = 2h(P) + 2h(Q) + O(1).$$

There are a lot of corollaries of this.

Corollary 6.5.1. $h(P + Q) \leq 2h(P) + C_{E,Q}$.

Proof. $h(P + Q) = 2h(P) + 2h(Q) + O_E(1) - h(P - Q)$, and we're done because $h(P - Q) \geq 0$. □

Corollary 6.5.2. $h(mP) \geq m^2 h(P) - C_E$.

Proof. We actually can prove that $h(mP) = m^2 h(P) + O(1)$. And we can prove this by induction using the parallelogram law with $Q = P$ and $P = mP$: we get

$$h\left((m + 1)P\right) = 2h(mP) + 2h(P) - h\left((m - 1)P\right) + O(1).$$

And we use $h(\mathcal{O}) = 0$ and $h(1 \cdot P) = h(P)$. □

In summary:

Theorem 6.5.3 (Mordell-Weil)**.** *$E(K)$ is finitely generated.*

6.6 Canonical (Néron-Tate) height

The idea: we just proved that $h(mP) = m^2 h(P) + O_E(m^2)$. So why not consider $\lim_{m\to\infty} m^{-2}h(mP)$ and get a unique formula? Well, the $O(1)$ depends on m, so this doesn't quite work a priori. But in fact, one can show that this quotient converges:

Theorem 6.6.1. *Fix $m \geq 2$. Then the limit*

$$\hat{h}(P) := \lim_{k\to\infty} \frac{1}{m^{2k}} h(m^k P)$$

exists, is independent of m, and satisfies:

(a) $\hat{h}(mP) = m^2 \hat{h}(P)$.

(b) $\hat{h} = h + O_E(1)$.

(c) $\hat{h}(P+Q) + \hat{h}(P-Q) = 2\hat{h}(P) + 2\hat{h}(Q)$

(d) $\hat{h}(P) \geq 0$, *and* $\hat{h}(P) = 0$ *if and only if* $P \in E_{tors}$.[14]

(e) Define a pairing

$$\langle \cdot, \cdot \rangle : E \times E \to \mathbb{R} : (P, Q) \mapsto \hat{h}(P+Q) - \hat{h}(P) - \hat{h}(Q).$$

Then:

(i) $\langle \cdot, \cdot \rangle$ *is a bilinear form.*

(ii) The associated quadratic form $P \mapsto \langle P, P \rangle$ is positive definite on $E(K)/E(K)_{tors}$, and in fact it's positive definite on $E(K) \otimes_{\mathbb{Z}} \mathbb{R}$.

Proof. We'll show the limit exists using telescoping sums. We already showed that $|h(mQ) - m^2 h(Q)| \leq C_{E,m}$, for every Q. We will show that $\frac{h(m^k P)}{m^{2k}}$ is a Cauchy sequence. Towards this, we can estimate

$$\begin{aligned}
\left| \frac{h(m^k P)}{m^{2k}} - \frac{h(m^\ell P)}{m^{2\ell}} \right| &= \left| \sum_{i=\ell}^{k-1} \left(\frac{h(m^{i+1}P)}{m^{2(i+1)}} - \frac{h(m^i P)}{m^{2i}} \right) \right| \\
&\leq \sum_{i=\ell}^{k-1} \frac{1}{m^{2i+2}} \left| h(m^{i+1}P) - m^2 h(m^i P) \right| \\
&\leq \sum_{i=\ell}^{\infty} \frac{1}{m^{2i+2}} C_{E,m} \\
&= \frac{C_{E,m}}{m^{2\ell}} \frac{1}{m^2 - 1},
\end{aligned}$$

which limits to zero as $k, \ell \to \infty$. Therefore the limit $\hat{h}(P)$ exists. Now let us put $\ell = 0$, so we get

$$\left| \frac{h(m^k P)}{m^{2k}} - h(P) \right| \leq C_{E,m},$$

so taking $m \to \infty$ gives that $|\hat{h}(P) - h(P)| \leq C_{E,m}$. Note that for all n, the canonical height is

$$\hat{h}_m(nP) = \lim_{k\to\infty} \frac{h(m^k nP)}{m^{2k}} = \lim_{k\to\infty} \frac{h(n \cdot m^k P)}{m^{2k}} = \lim_{k\to\infty} \frac{n^2 h(m^k P) + O_{n,E}}{m^{2k}} = n^2 \hat{h}_m(P),$$

and from there it's easy to see that the m-canonical height and the n-canonical height are the same, as

$$\hat{h}_m(P) = \frac{1}{m^{2r}} \hat{h}_m(m^r P),$$

[14] This is the elliptic curve analog of Kronecker's theorem; it's due to Néron and Tate.

and on the other hand,

$$h(n^r P) + O_m(1) = \hat{h}_m(n^r P) = n^r \hat{h}_m(P),$$

so if we divide by n^r then we get

$$\frac{h(n^r P)}{n^r} + \frac{O_m(1)}{n^r} = \hat{h}_m(P),$$

and if we take $r \to \infty$ then the LHS becomes $\hat{h}_n(P)$, as needed. For (c), we can compute that

$$h(m^k(P+Q)) + h(m^k(P-Q)) = 2h(m^k P) + 2h(m^k Q) + O_E(1),$$

so we divide both sides by m^{2k} and take $k \to \infty$. For (d), if $P \in E_{tors}$, then it has finite order so $mP = \mathcal{O}$, and $\hat{h}(\mathcal{O}) = \hat{h}(mP) = m^2 \hat{h}(P)$, so we can divide by m^2. The other direction is less of a triviality; if $\hat{h}(P) = 0$, then $m^2 \hat{h}(P) = 0$ for all m, which implies that $\hat{h}(mP) = 0$ for all m by the transformation law for the canonical height; this implies that $\hat{h} = h + O_E(1)$, so on the one hand, $|h(mP) - \hat{h}(mP)| \le C_E$, but this is equivalent to $|h(mP)| \le C_E$. This implies that $\{mP : m \in \mathbb{Z}\} \subseteq \{Q \in E(K) : h(Q) \le C_E\}$. But this latter set is finite. This implies there exists $n > m$ with $nP = mP$, by the pigeonhole principle, so $(n-m)P = \mathcal{O}$ implies P is a torsion point.

For (e), a standard exercise: if $F(x, y)$ satisfies the parallelogram law on an abelian group, then the associated pairing is bilinear. One needs to show that $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$, and what this reduces to is an identity on three variables that looks like inclusion-exclusion.

(Lecture 33: April 7, 2021)

It's clear that $\hat{h}$ is a positive-definite quadratic form on $E(K)/E(K)_{tors}$, because this is $\mathbb{Z}^{\operatorname{rank} E(K)}$. Today we'll argue that $\hat{h}$ is a positive-definite quadratic form $E(K) \otimes_{\mathbb{Z}} \mathbb{R} \to \mathbb{R}$. Notice that $E(K) \otimes_{\mathbb{Z}} \mathbb{R} \cong \mathbb{R}^{\operatorname{rank} E(K)}$.

Example 6.6.2. Consider the map $q : \mathbb{Z}^2 \to \mathbb{R}$ defined by $q(x, y) = |x - \sqrt{2}y|^2$. This is a quadratic form, it satisfies $q \ge 0$, and $q = 0$ if and only if $(x, y) = (0, 0)$. But it's no longer positive definite on $\mathbb{Z}^2 \otimes \mathbb{R} = (\mathbb{Z} \otimes \mathbb{R}) \oplus (\mathbb{Z} \otimes \mathbb{R})$, because $q(1 \otimes \sqrt{2}, 1 \otimes 1) = 0$.

We know $\hat{h} \ge 0$, and $\hat{h}(P) = 0$ if and only if $P = 0$ for $P \in E(K) \otimes \mathbb{Q}$. We also know that $\{P \in E(K)/E(K)_{tors} : \hat{h}(P) \le C\}$ is finite. How should we think of this? $E(K)/E(K)_{tors} \cong \mathbb{Z}^r$ injects into $E(K) \otimes \mathbb{R} \cong \mathbb{R}^r$, so we have a lattice in this real vector space, and we have a quadratic form $\hat{h}$. We can use the quadratic form to define distances, angles, and volumes on this lattice such that these attributes contain arithmetic information.

Proposition 6.6.3. *Suppose $L \subseteq V$ is a lattice inside a finite dimensional real vector space. Suppose $q : V \to \mathbb{R}$ is a positive semi-definite quadratic form. Assume that:*

(a) For every $P \in L$, we have $q(P) = 0$ if and only if $P = 0$.

(b) $\{P \in L : q(P) \le C\}$ is finite for all C.

Then q is positive definite on V, i.e. for every $P \in V$, $q(P) = 0$ if and only if $P = 0$.

Proof of proposition. We can find an $\mathbb{R}$-basis $v_1, \dots, v_d$ for V so that q is diagonal with 1's and then -1's and then 0's descending along the diagonal. In other words, $q(x) = \sum_{i=1}^{s} x_i^2 - \sum_{i=s+1}^{s+t} x_i^2$. We will use this basis to identify $V \cong \mathbb{R}^d$, and let μ be the Lebesgue measure on $\mathbb{R}^d$. Recall that Minkowski's theorem says the following: let $B \subseteq V$; if $\mu(B) \ge C(V, L, q)$, and B is convex and symmetric, then $B \cap L$ contains a nonzero point. In our case, for $\epsilon, \delta > 0$, look at

$$B(\epsilon, \delta) := \{x \in V : \sum_{i=1}^{s} x_i^2 \le \epsilon, \sum_{i=s+1}^{s+t} x_i^2 \le \delta\}.$$

This is convex as it's the intersection of two hyperspheres, and it's symmetric. And

$$\operatorname{Vol} B(\epsilon, \delta) = \begin{cases} \infty & s + t < d \\ (\text{volume of unit sphere}) \cdot \epsilon^s \delta^t & s + t = d. \end{cases}$$

Let $\lambda = \inf\{q(P) : P \in L, P \neq 0\}$. Property (b) tells us that $\lambda > 0$. Towards a contradiction, suppose q is not positive definite on V. Recall that $q(x) = x_1^2 + \cdots + x_s^2 - x_{s-1}^2 - \cdots - x_{s+t}^2$, hence it must be the case that $s < d$. Then the shape $B(\lambda/2, \text{big } \delta)$ satisfies

$$\mu B(\lambda/2, \text{big } \delta) \begin{cases} = \infty \\ \gg \epsilon^t. \end{cases}$$

Therefore there exists δ so that $\mu B(\lambda/2, \delta)$ is greater than the Minkowski constant; this implies that there exists a lattice point $P \neq 0$ with $P \in B(\lambda/2, \delta)$. But we can compute that $q(P) \leq \lambda/2$, which is a contradiction, as λ is the smallest nonzero q-value for $P \in L$. This completes the proof of the proposition. □

That proposition lets us deduce that $\hat{h}$ is a positive definite quadratic form on $E(K) \otimes \mathbb{R}$. This completes the proof. □

To recapitulate, we have $E(K) \otimes \mathbb{R}$ a finite dimensional vector space, $\hat{h}$ a positive definite quadratic form on this vector space, and $E(K)/E(K)_{tors}$ a lattice in this vector space, and this lattice is the image of $E(K) \to E(K) \otimes \mathbb{R}$. This is a familiar setup: the ring of integers $\mathcal{O}_K$ sits as a lattice inside $\mathcal{O}_K \otimes_{\mathbb{Q}} \mathbb{R}$, and the group $\mathcal{O}_K^*/\mu$ sits as a lattice inside $\mathbb{R}^{r_1+r_2-1}$. Associated to these lattices we have the discriminant and regulator. In our case:

Definition 6.6.4. Let $P_1, \ldots, P_r \in E(K)$ be a basis for $E(K)/E(K)_{tors}$. The *elliptic regulator* of E/K is

$$\text{Reg}(E/K) := \det(\langle P_i, P_j \rangle)_{i \leq i, j \leq r},$$

where $\langle P_i, P_j \rangle$ is given by the canonical $\hat{h}$ pairing. In fact, this is equal to the volume of the fundamental domain of $E(K)/E(K)_{tors}$ in $E(K) \otimes \mathbb{R}$, where $E(K) \otimes \mathbb{R}$ is a Euclidian space whose norm is given by $\hat{h}$.

Thus, $\text{Reg}(E/K)$ measures the arithmetic complexity of the "free part" of $E(K)$. If $\operatorname{rank} E(\mathbb{Q}) = r$, then the Birch Swinnerton Dyer conjecture over $\mathbb{Q}$ says that $L(E/\mathbb{Q}, s) = C(s-1)^r + \text{higher order terms}$, where $C = c\frac{\text{Reg}(E/\mathbb{Q})}{|E(\mathbb{Q})_{tors}|^2}$. The BSWD conjecture over $\mathbb{Q}$ implies that there exists an effective algorithm to compute $E(\mathbb{Q})$. This observation was originally made by Manin.

We know $h(P)$ measures the complexity of P, and because $h(P)$ and $\hat{h}(P)$ differ by $O(1)$, we have that $\hat{h}(P)$ also measures the complexity of P. Because $\hat{h}(P) = 0$ if and only if $P \in E_{tors}$, this begs the question: if $P \notin E_{tors}$, how small can $\hat{h}(P)$ be? What do we mean by "how small"? The idea is that if E is complicated, its points should be complicated. Formally:

Definition 6.6.5. Take an elliptic curve $E/\mathbb{Q}$, and take a Weierstrass equation $y^2 = x^3 = Ax + B$ with A, B integers such that $\gcd(A^3, B^2)$ is 12'th power free. Define the *height* of E by

$$h(E) := \log \max\{|A|^3, |B|^2\}.$$

Conjecture 6.6.6 (Lang's height conjecture)**.** *There exist absolute constants $c_1, c_2 > 0$ so that for all elliptic curves $E/\mathbb{Q}$ and all non-torsion points $P \in E(\mathbb{Q})$,*

$$\hat{h}(P) \geq c_1 h(E) - c_2.$$

That's an open problem, and in fact, the ABC conjecture implies Lang's height conjecture.

Conjecture 6.6.7 (ABC)**.** *Suppose $a, b, c \in \mathbb{Z}$, and $a + b = c$, and $\gcd(a, b, c) = 1$. Then*

$$\max\{|a|, |b|, |c|\} \leq K_\epsilon \prod_{p|abc} p^{1+\epsilon}.$$

(Lecture 34: April 12, 2021)

6.7 Group cohomology: how we study Mordell-Weil groups in a modern setting

Let M be an abelian group (thought of as a $\mathbb{Z}$-module) and let G be a profinite group acting continuously on M.

Example 6.7.1. Let $G = G_K = \mathrm{Gal}(\overline{K}/K)$, and $M = \overline{K}^*$ or $M = E(\overline{K})$.

We look at the functor

$$M \to M^G := \{m \in M : \sigma(m) = m,\ \forall \sigma \in G\}.$$

Example 6.7.2. $(\overline{K}^*)^{G_K} = K^*$ by the fundamental theorem of Galois theory, and $E(\overline{K})^{G_K} = E(K)$.

Definition 6.7.3. The *first cohomology group* is $H^0(G_K, M) := M^G$.

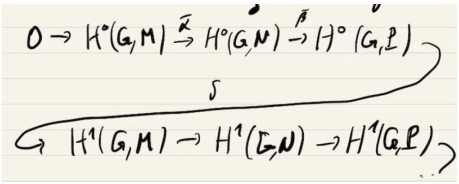
Proposition 6.7.4. *Suppose we have an exact sequence*

$$0 \xrightarrow{\alpha} M \xrightarrow{\beta} N \to P \to 0$$

of G-modules. Then there is an exact sequence

$$0 \to H^0(G, M) \xrightarrow{\overline{\alpha}} H^0(G, N) \xrightarrow{\overline{\beta}} H^0(G, P).$$

In other words, this functor is left-exact, but not right exact. But in fact, we get a long exact sequence



where

$$H^1(G, M) := \frac{\{\textit{maps } \xi : G \to M : \xi_{\sigma\tau} = \xi_\sigma + \sigma(\xi_\tau) \textit{ for all } \sigma, \tau \in G\}}{\{\textit{maps } G \to M \textit{ of the form } \sigma \to \sigma(m) - m \textit{ for some } m \in M\}}.$$

Example 6.7.5. Take $\overline{K}^* \to \overline{K}^* : x \mapsto x^m$. This is onto because we're over $\overline{K}$, and the kernel is roots of unity, so this gives the exact sequence

$$1 \to \mu_m \to \overline{K}^* \to \overline{K}^* \to 1.$$

The numerator set is called "1-cocycles" and the denominator is "1-coboundaries." Note: if the action of G on M is trivial, i.e. $\sigma(m) = m$ for all σ and m, then the cocycles are homomorphisms, and the coboundaries are zero maps, so in this case $H^1(G, M) = \mathrm{Hom}(G, M)$. And this is an interesting nontrivial group.

What is the connecting homomorphism $H^0(G, P) \xrightarrow{\delta} H^1(G, M)$, i.e. $P^G \xrightarrow{\delta} H^1(G, M)$? As $N \to P$ is onto, given $p \in P^G$, we can choose some $n \in N$ with $\beta(n) = p$. Is n fixed by G? This is the case only if $G \to N : \sigma \mapsto \sigma(n) - n$ is the zero map. Note that

$$\beta(\sigma(n) - n) = \beta(\sigma(n)) - \beta(n) = \sigma(\beta(n)) - \beta(n) = \sigma(p) - p = p - p = 0,$$

so $\sigma(n) - n \in \ker \beta = \mathrm{Image}(\alpha)$. But α is injective, so there exists a unique $m_{n,p,\sigma} \in M$ with $\alpha(m_{n,p,\sigma}) = \sigma(n) - n$. This gives a map $G \to M : \sigma \mapsto m_{n,p,\sigma}$ which satisfies $\sigma(m_{n,p,\sigma}) = \sigma(n) - n$. One can check that this map $G \to M$ satisfies this 1-cocycle condition. If we choose some other n' with $\beta(n') = p$, then $\sigma \mapsto m_{n,p,\sigma} - m_{n',p,\sigma}$ is in fact a 1-coboundary for $G \to M$, so it's a trivial element of the cohomology group. In summary, we've just defined a map

$$H^0(G, P) \to H^1(G, M) : P \mapsto [\sigma \mapsto m_{n,p,\sigma}],$$

where $n \in N$ is any element which satisfies $\beta(n) = p$.

We will need the "Inflation-Restriction Sequence." The setting for this: consider a G-module M, and $H \subseteq G$ a normal subgroup. Then M is an H-module, so $M^G \subseteq M^H$. As M^H is a G/H-module, there is an exact sequence

$$0 \to H^1(G/H, M^H) \xrightarrow{\text{inflation}} H^1(G, M) \xrightarrow{\text{restriction}} H^1(H, M).$$

This lets us break the group G up into the pieces H and G/H. Note that this inflation-restriction sequence is actually part of a spectral sequence, due to Serre; the relevant cohomology groups are

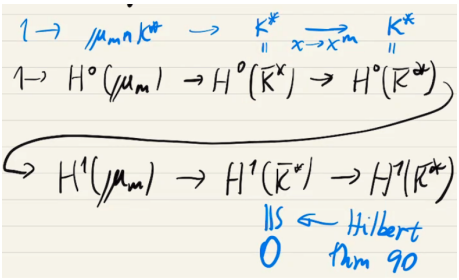
$$H^q(G/H, H^p(H, M)) \implies H^{p+q}(G, M),$$

which converges in the context of spectral sequences.

Example 6.7.6. Consider the short exact sequence

$$1 \to \mu_m \to \overline{K}^* \xrightarrow{x \mapsto x^m} \overline{K}^* \to 1$$

of G_K-modules. Then we have the long exact sequence



This gives the *Kummer isomorphism*

$$K^*/(K^*)^m \xrightarrow{\sim} H^1(G_K, \mu_m) : b \mapsto \left[\sigma \mapsto \frac{\sigma(\sqrt[m]{b})}{\sqrt[m]{b}}\right].$$

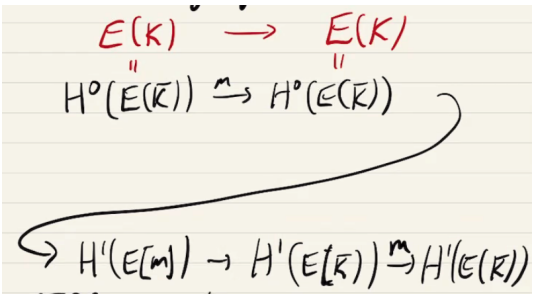
This isomorphism actually contains the main theorem of Kummer theory. How is this? If $\mu_m \subseteq K$, then Galois acts trivially on μ_m, so $H^1(G_K, \mu_m) \cong \mathrm{Hom}(G_K, \mu_m)$. But this latter group corresponds with cyclic Galois extensions, $\{\text{Galois } L/K : G_{L/K} \subseteq \mathbb{Z}/m\mathbb{Z}\}$, via the correspondence $(\phi : G_K \to \mu_m) \mapsto \overline{K}^{\ker \phi}$, as $\mathrm{Gal}(\overline{K}^{\ker \phi}/K) \hookrightarrow \mu_m$ by Galois theory. The corollary of this is the main theorem of Kummer theory:

Corollary 6.7.7. *If $\mu_m \subseteq K$, and if L/K is abelian with $G_{L/K} \subseteq \mathbb{Z}/m\mathbb{Z}$, then $L = K(\sqrt[m]{b})$ for some $b \in K^*$.*

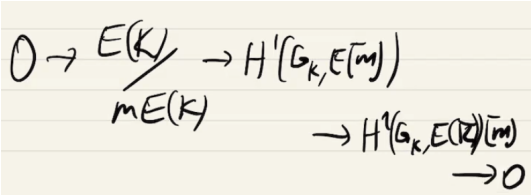
One can take the Kummer sequence for the elliptic curve,

$$0 \to E[m] \to E(\overline{K}) \xrightarrow{m} E(\overline{K}) \to 0.$$

This is an exact sequence of G_K-modules (i.e. an exact sequence of abelian groups on which G_K acts). Then we get the long exact sequence



which gives the short exact *elliptic Kummer sequence*



This is a useful tool for the study of elliptic curves.

6.8 Exercises

Exercise (Silverman 8.7(b)). Let

$$v_K(N, C) := \{P \in \mathbb{P}^N(K) : H_K(P) \leq C\}.$$

Prove that

$$\lim_{C \to \infty} \frac{v_{\mathbb{Q}}(N, C)}{C^{N+1}} = \frac{2^N}{\zeta(N+1)}.$$

Proof. By definition of the height function, it suffices to show that

$$\#\{(x_0, \ldots, x_N) \in \mathbb{N}^{N+1} : |x_0|, \ldots, |x_N| \leq C, \gcd(x_0, \ldots, x_N) = 1\} = \frac{C^{N+1}}{\zeta(N+1)} + o(C^{N+1})$$

as $C \to \infty$, because the factor of 2^N adds all possible sign combinations lists $v_{\mathbb{Q}}(N, C)$. Notice that $(1 - p^{-(N+1)})$ is the probability that no $N+1$ independently chosen natural numbers are divisible by the prime p, so the probability that no $N+1$ independently chosen natural numbers are divisible by *any* prime is

$$\prod_p (1 - p^{-N-1}) = \prod_p \left(\frac{1}{1 - p^{-(N+1)}}\right)^{-1} = \frac{1}{\zeta(N+1)}.$$

This completes the proof. □

Exercise (Silverman 8.8). Prove the following basic properties of height functions.

(a) For any $x_1, \ldots, x_N \in \overline{\mathbb{Q}}$, we have $H(x_1 \cdots x_N) \leq H(x_1) \cdots H(x_N)$.

(b) For any $x_1, \ldots, x_N \in \overline{\mathbb{Q}}$, we have $H(x_1 + \cdots + x_N) \leq N H(x_1) \cdots H(x_N)$.

Proof. Choose K so that every $x_i \in K$. For (a), we can directly estimate that

$$\begin{aligned} H_K(x_1 \cdots x_N) &= \prod_{v \in M_K} \max\{|x_1 \cdots x_N|_v, 1\}^{n_v} \\ &\leq \prod_{v \in M_K} (\max\{|x_1|_v, 1\} \cdots \max\{|x_N|_v, 1\})^{n_v} \\ &= H_K(x_1) \cdots H_K(x_N). \end{aligned}$$

And for (b), we can compute that

$$\begin{aligned} H_K(x_1+\cdots+x_N) &= \prod_{v\in M_K} \max\{|x_1+\cdots+x_N|_v,1\}^{n_v} \\ &\le \prod_{v\in M_K} \max\{N^{\epsilon_v}\max\{|x_1|_v,\ldots,|x_N|_v\},1\}^{n_v} \\ &\le \prod_{v\in M_K} N^{\epsilon_v n_v}\left(\max\{|x_1|_v,1\}\cdots\max\{|x_N|_v,1\}\right)^{n_v} \\ &= \left(\prod_{v\in M_K} N^{\epsilon_v n_v}\right) H_K(x_1)\cdots H_K(x_N). \end{aligned}$$

If we take the $[K:\mathbb{Q}]$'th root of this estimate, the result now follows from the bound $0 \le \epsilon_v n_v \le 1\cdot n_v \le [K:\mathbb{Q}]$. □

Exercise (Silverman 8.10). Let F be the rational map

$$F : \mathbb{P}^2 \to \mathbb{P}^2 : [x,y,z] \mapsto [x^2, xy, z^2].$$

Note that F is a morphism at every point except at $[0,1,0]$, where it is not defined. Prove that there are infinitely many points $P \in \mathbb{P}^2(\mathbb{Q})$ such that

$$H(F(P)) = H(P).$$

Proof. We can compute that $F([1,y,1]) = [1,y,1]$, which implies that F has infinitely many fixed points. □

Exercise (Silverman 8.12). Calculate $E(\mathbb{Q})_{tors}$ for each of the following elliptic curves:

(1) $y^2 = x^3 - 2$
(2) $y^2 = x^3 + 8$
(3) $y^2 = x^3 + 4$
(4) $y^2 = x^3 + 4x$
(5) $y^2 - y = x^3 - x^2$
(6) $y^2 = x^3 + 1$
(7) $y^2 = x^3 - 43x + 166$
(8) $y^2 + 7xy = x^3 + 16x$
(9) $y^2 + xy + y = x^3 - x^2 - 14x + 29$
(10) $y^2 + xy = x^3 - 45x + 81$
(11) $y^2 + 43xy - 210y = x^3 - 210x^2$
(12) $y^2 = x^3 - 4x$
(13) $y^2 = x^3 + 2x^2 - 3x$
(14) $y^2 + 5xy - 6y = x^3 - 3x^2$
(15) $y^2 + 17xy - 120y = x^3 - 60x^2$

Solution. The following Sage code outputs the list of torsion subgroups:

```
coeff_lists = [[0,0,0,0,-2], [0,0,0,0,8], [0,0,0,0,4], [0,0,0,4,0],
[0,-1,-1,0,0], [0,0,0,0,1], [0,0,0,-43,166], [7,0,0,16,0], [1,-1,1,-14,29],
[1,0,0,-45,81], [43,-210,-210,0,0], [0,0,0,-4,0], [0,2,0,-3,0],
[5,-3,-6,0,0], [17,-60,-120,0,0]]

for coeffs in coeff_lists:
        E = EllipticCurve(coeffs)
        T = E.torsion_subgroup()
        print(T)
```

It tells us that for $n = 1,\ldots,10$, the elliptic curve in position (n) has torsion subgroup $\mathbb{Z}/n\mathbb{Z}$, and the remaining five torsion subgroups are $\mathbb{Z}/12\mathbb{Z}$, $\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/2\mathbb{Z}$, $\mathbb{Z}/4\mathbb{Z}\oplus\mathbb{Z}/2\mathbb{Z}$, $\mathbb{Z}/6\mathbb{Z}\oplus\mathbb{Z}/2\mathbb{Z}$, and $\mathbb{Z}/8\mathbb{Z}\oplus\mathbb{Z}/2\mathbb{Z}$, respectively. □

Exercise (Silverman 8.16). Let V be a finite-dimensional real vector space and let $L \subseteq V$ be a lattice, i.e. L is a discrete subgroup of V containing a basis for V. Let $q : V \to \mathbb{R}$ be a quadratic form such that $q(P) = 0$ on L if and only if $P = 0$, but $\{P \in V : q(P) \leq C\}$ is not finite for every C. Show that under these weaker hypotheses, we can't necessarily conclude that q is positive definite on V.

Proof. Consider the quadratic form

$$q : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto (x - \sqrt{2}y)^2 = x^2 - 2\sqrt{2}xy + 2y^2,$$

as well as the lattice $\mathbb{Z}^2 \subseteq \mathbb{R}^2$. Then $q(x, y) = 0$ if and only if $x - 2\sqrt{y} = 0$, and because 2 is irrational, this equation has only the trivial solution over $\mathbb{Z}$. However, q is not positive definite on V, because $q(\sqrt{2}, 1) = 0$. But by density of $\mathbb{Q} \in \mathbb{R}$, we know $\{P \in L : q(P) < \epsilon\}$ is infinite for every $\epsilon > 0$, as integral solutions to $q(x, y) < \epsilon$ as $\epsilon \to 0$ correspond to successively better rational approximations of $\sqrt{2}$. □