TORSION ON ELLIPTIC CURVES AND MAZUR'S THEOREM

SPENCER DEMBNER

Abstract. We discuss several results about rational torsion points on elliptic curves. First, we outline the proof of the Mordell-Weil Theorem, a fundamental result which ensures the subgroup is finite. Next, we explain how to compute the torsion of a specific elliptic curve using local methods. Finally, we discuss modular curves, which make it possible to study torsion points of every elliptic curve at once. This allows us to state (but not prove) Mazur's celebrated torsion theorem, which says that only 15 specific groups can occur as rational torsion subgroups of elliptic curves.

Contents

1. Introduction

Our focus in this paper will be on elliptic curves, which are smooth, projective algebraic curves of genus one over a field. The fundamental fact about elliptic curves is that they admit a group law: given an elliptic curve E, there is a map of algebraic varieties $+\colon E \times E \to E$ which gives E the structure of an Abelian group. The group operation makes sense for any field, so for instance if the elliptic curve E is defined over $\mathbb{Q}$, the points on E with complex coordinates form a group, and the points with rational coordinates form a subgroup.

In 1922, Louis Mordell proved that given any elliptic curve E defined over $\mathbb{Q}$, its group $E(\mathbb{Q})$ of rational points is a finitely generated abelian group. In 1928, Andre Weil proved the *Mordell-Weil Theorem*, which shows that the same holds for any finite extension K of $\mathbb{Q}$. By the classification of finitely generated Abelian groups,

Date: August 19, 2019.

we find that $E(\mathbb{Q}) \cong \mathbb{Z}^r \times F$, where F is some finite Abelian group and r is the *rank* of the group $E(\mathbb{Q})$, also known as the rank of the curve E.

This immediately poses two additional questions: first, what are the possible ranks r? Second, what can we say about the finite torsion subgroup F? The first question turns out to be very difficult, and little is known: in fact, it is unknown whether there are elliptic curves whose rank grows arbitrarily large. The Birch and Swinnerton-Dyer Conjectures, still unproven, would establish a connection between the ranks of elliptic curves and certain L-functions.

The second question is much more tractable. In 1978, Barry Mazur proved the following elegant result, which completely settles it, at least over $\mathbb{Q}$:

Theorem 1.1. *Let E be an elliptic curve defined over $\mathbb{Q}$. The possible torsion subgroups of $E(\mathbb{Q})$ are:*

(1) $\mathbb{Z}/N\mathbb{Z}$, where $1 \leq N \leq 10$, or $N = 12$.
(2) $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2N\mathbb{Z}$, where $N = 1, 2, 3, 4$.[1]

In this paper, we develop some of the general theory of torsion on elliptic curves in order to put Mazur's Theorem in context. In the first few sections, we state fundamental results on elliptic curves, skipping all but the simplest proofs. In the next few sections, we outline the proof of Mordell-Weil in some detail, paying special attention to local considerations and the role of Galois cohomology. Then, in order to make the argument in the proof of Mordell-Weil more explicit, we discuss how to calculate the group of rational points on a specific curve, as well as how these methods can fail. In the final sections, we develop the basic theory of modular curves. These are (roughly) curves whose points correspond to elliptic curves with specified torsion points. Studying them provides an effective way to answer questions about torsion across all elliptic curves, and they played a central role in the proof of Mazur's Theorem. The proof of the full theorem is far beyond our scope, but we discuss proofs of several specific cases.

In general, we work over an arbitrary number field for the proof of Mordell-Weil. For the rest of the paper, we specialize to elliptic curves defined over $\mathbb{Q}$, where finding possible torsion subgroups is much easier. In the final section, we briefly consider elliptic curves defined over $\mathbb{C}$ rather than a number field.

In general, we assume the reader has seen the basic theory of algebraic curves, at the level of [4] or the first two chapters of [16]; in particular, a couple of results which we quote without proof depend on the Riemann-Roch theorem for curves. We also assume a solid knowledge of Galois theory, and familiarity with some basic notions from algebraic number theory. In the final sections, we will occasionally need basic facts from complex analysis, and these sections use very little from the rest of the paper.

We don't assume any prior exposure to elliptic curves, although some basic results will be quoted without proof. Most of these results come from [16], and can also be found in [10].

[1]The fact that every torsion subgroup is either of the form $\mathbb{Z}/N\mathbb{Z}$ or $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ follows from the structure of $E(\mathbb{R})$, which will either be $\mathbb{S}^1$ or $\mathbb{S}^1 \times \mathbb{Z}/2\mathbb{Z}$, depending on the discriminant. For a discussion, see Chapter V of [15].

2. Basic Theory

For the purposes of this paper, all curves considered will be projective, unless stated otherwise. We will need a couple of basic results from the theory of algebraic curves, which we state here for reference.

Theorem 2.1 (Bézout's Theorem)**.** *: Let C, C' be projective plane curves of degree m and n respectively. Then C, C' intersect with total multiplicity mn.*

Proof. See Chapter 5 of [4]. □

Theorem 2.2 (Degree-genus formula)**.** *Let C be a smooth projective plane curve defined by an equation of degree d, with genus g. Suppose further that C is defined over a field of characteristic 0. Then:*

$$g = \frac{(d-1)(d-2)}{2}$$

Proof. This is a consequence of the *Riemann-Hurwitz formula*, which states that if $\phi\colon C_1 \to C_2$ is a degree n map between smooth curves C_1, C_2 of genus g_1, g_2 respectively, both defined over a field K with $\mathrm{char}(K) = 0$, then:

$$2g_1 - 2 = n(2g_2 - 2) + \sum_{P \in C_1} (e_\phi(P) - 1)$$

Here, $e_\phi(P)$ is the ramification index of ϕ at P, which is one for all but finitely many P. For a proof of this formula, see Chapter II of [16]. In order to prove the degree-genus formula, pick a line $L \subset \mathbb{P}^2$ which is not tangent to C, and let $\phi\colon C \to L$ be the projection map. The line L is isomorphic to $\mathbb{P}^1$, and hence has genus 0 (the genus of $\mathbb{P}^1$ can be calculated from the Riemann-Roch theorem, discussed in Chapter 8 of [4]). The map ϕ will have degree d, by Bézout's Theorem. Likewise, by Bézout's Theorem, since L is not tangent to C, $L \cap C$ will consist of d distinct points, which are the ramification points of the map, and each such point will ramify to degree d. Thus, we have:

$$2g(C) - 2 = -2d + d(d-1)$$

Rearranging terms, we find that $g(C) = \frac{(d-1)(d-2)}{2}$, as desired. □

First, we give the general definition of an elliptic curve:

Definition 2.3. Let K be any field. An elliptic curve defined over K is a pair (E, O), where:

(1) E is a projective algebraic curve of genus one, defined over K.
(2) O is a point on E with coordinates in K (a 'K-rational point').

Note that the point O is part of the definition. In particular, if a curve has an equation with coefficients in K but has no K-rational points, then the curve is not an elliptic curve defined over K. When we give the curve E a group law, O will be its identity element. For ease of notation, however, we will usually just call a curve E, allowing the point O to be understood. We will also write E/K to indicate that E is defined over the field K.

In order to simplify proofs, it helps to introduce a more explicit description of elliptic curves. This motivates the following definition:

Definition 2.4. A Weierstrass equation is an equation of the form:

$$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6 \tag{2.5}$$

where a_1, a_2, a_3, a_4, a_6 are contained in some field K. Working in homogenous projective coordinates, this corresponds to the equation:

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$$

Every Weierstrass equation defines a plane curve, C. The equation has an associated number Δ, called the discriminant, whose definition we give in a special case below. The discriminant Δ is nonzero if, and only if, the associated curve C is nonsingular (for more details on this, see Chapter III of [16]). Note that if C is nonsingular, then C can be made into an elliptic curve defined over K. Indeed, by the degree-genus formula, C will have genus one, so we only need to check that C contains a K-rational point O. For this, note that the point $[0 : 1 : 0]$ is K-rational, and is the unique point at infinity contained on the curve C. By convention, if a curve C is defined by a Weierstrass equation, we always take the point O to be the point at infinity.

What is less clear is that the converse is true. That is, an elliptic curve can always be written in Weierstrass form:

Lemma 2.6. *Let E/K be an elliptic curve. There are functions $x, y \in K(E)$ such that the map $\phi\colon E \to \mathbb{P}^2$ given by:*

$$\phi(P) = [x(P), y(P), 1]$$

is an isomorphism.

Proof. See [16], Proposition III.3.1. □

In particular, we now know that every elliptic curve is a plane curve.

Over a field of characteristic 0, such as $\mathbb{Q}$, we can change coordinates to write any Weierstrass equation in the simpler form:

$$y^2 = x^3 + Ax + B$$

We will usually work with Weierstrass equations of this form. In this case, the discriminant is given by:

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

We also associate another quantity, called the j-invariant, and defined by:

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

The j-invariant will become important in the final section when we study modular curves. This is because of the following fact:

Fact 2.7. Suppose K is algebraically closed. Then two elliptic curve E, E' are isomorphic over K if and only if they have the same j-invariant.

Elliptic curves have a group law which is compatible with their algebraic structure. More formally, we have:

Theorem 2.8. *Let (E, O) be an elliptic curve. Then there exists a map $+ : E \times E \to E$, such that:*

1. *E is an Abelian group with the binary operation $+$, with identity element O.*

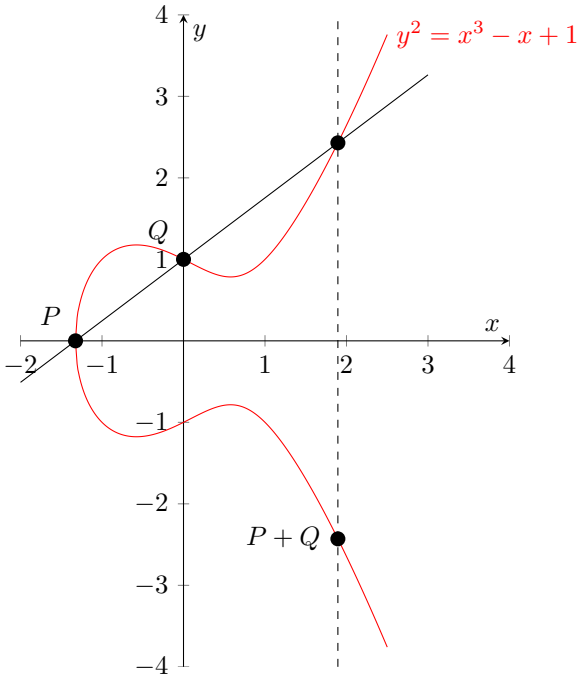


FIGURE 1. An illustration of the group law on the elliptic curve $y^2 = x^3 - x + 1$

(2) The map $+$ is a morphism of varieties.

Proof. By the above discussion, we may take E to be a plane curve, allowing us to apply Bézout's Theorem. The group law on E is given by the following algorithm: given two points $A, B \in E$, the line between A and B intersects E at a unique third point P, by Bézout's Theorem (if $A = B$, take the tangent line). The line between P and O likewise intersects at a unique third point P', and we define $A + B = P'$. A proof that this gives an Abelian group law, via the Riemann-Roch theorem, is given in Chapter III of [16]. For a different argument, see Chapter 5 of [4]. □

We now define maps between elliptic curves. An elliptic curve is compatibly both a group and an algebraic variety, so it makes sense to require that maps preserve both structures. Thus, we define an *isogeny* to be a map $f\colon E \to E'$ between elliptic curves which is a morphism of varieties and also a group homomorphism. It turns out that a weaker condition is enough:

Proposition 2.9. *Suppose $f\colon E \to E'$ is a morphism of varieties and maps $O \in E$ to $O' \in E'$. Then f is an isogeny.*

Proof. See [16], Theorem III.4.8. □

Every elliptic curve E has corresponding isogenies $[m]\colon E \to E$, mapping P to mP. Many basic properties of elliptic curves can be verified by carefully considering these maps. Here is one example:

Lemma 2.10. *Let E/K be an elliptic curve, pick $m \in \mathbb{Z}^{>0}$, let $E[m]$ be the subgroup of $E(K)$ consisting of m-torsion points, and suppose K is algebraically closed of characteristic 0. Then $E[m] \cong (\mathbb{Z}/m\mathbb{Z})^2$.*

Proof. The kernel of the isogeny $[m]$ has size m^2 for all m ([16], Theorem III.6.2), and this kernel is exactly $E[m]$. If n divides m, then $E[n] \subset E[m]$. Writing $G = E[m]$, this means that for every n dividing m, the n-torsion subgroup of G has size n^2. But the only Abelian group of size m^2 with this property is $(\mathbb{Z}/m\mathbb{Z})^2$. □

Remark 2.11. A version of this statement is true in characteristic p, but we won't need it. See [16], Corollary III.6.4.

3. Local Fields, and a Result on Torsion

For the proof of the Mordell-Weil Theorem, we will need several basic facts about local fields and elliptic curves defined over them. We state them here for reference. For proofs of the basic facts on local fields and definitions of terminology (absolute value, completion, $\mathbb{Q}_p$ local field, and so on), see any standard book on algebraic number theory such as [12] or [14]

Let K be a finite extension of $\mathbb{Q}$ (a *number field*). Given an absolute value v on K (which we always assume to be nontrivial), we will often want to consider the completion K_v of v with respect to K. If v is Archimedean, K_v will be isomorphic either to $\mathbb{R}$ or to $\mathbb{C}$. Otherwise. K_v will be isomorphic to a finite extension of $\mathbb{Q}_p$ for some prime p, and hence will be a local field. We will often make use of the following result, which characterizes the set of absolute values on K.

Theorem 3.1 (Ostrowski)**.** *Let K be a number field.*

1. *The only fields complete with respect to an Archimedean absolute value are $\mathbb{R}$ and $\mathbb{C}$. In particular, the Archimedean absolute values on K correspond to real embeddings of K or conjugate pairs of complex embeddings.*
2. *The non-Archimedean absolute values on K are exactly the $\mathfrak{p}$-adic absolute values, where $\mathfrak{p}$ is a prime ideal in the ring of integers $\mathcal{O}_K$. The completions with respect to non-Archimedean absolute values are finite extensions of $\mathbb{Q}_p$ for some prime p.*

Proof. See [2]. □

Let K be a local field. The absolute value on K corresponds to a unique normalized discrete valuation, mapping elements of K to $\mathbb{Z} \cup \{\infty\}$; this is because K is the fraction field of the ring R of elements with absolute value at most 1, and the ring R is a discrete valuation ring. We always assume discrete valuations are normalized, which means that the minimum positive valuation of any element is 1. We will need the fact that valuations extend to algebraic extensions, and that finite extensions of local fields are local.

Lemma 3.2. *Let K be local. If E/K is an algebraic extension, the discrete valuation on K extends uniquely to a valuation on E. If E is assumed to be finite, then the valuation on E will be discrete, making E into a local field as well.*

Proof. See Chapter II, §3 of [14]. □

Let E/K be a finite extension of local fields. Suppose that $\pi \in K$ is an element of valuation 1 in K. We define the ramification degree of the extension, necessarily a positive integer, to be the valuation of π in E.

Now, suppose that L/K is a *Galois* extension of number fields, not necessarily finite, and let v be a valuation on K. The valuation v may have multiple extensions to L, since K is not complete with respect to v. It turns out that $\mathrm{Gal}(L/K)$ acts on the set of valuations extending v by sending the valuation w to $w \circ \sigma$. For fixed w, we define the decomposition group $G_w \subset \mathrm{Gal}(L/K)$ to be the stabilizer of the valuation w. In general, not all the decomposition groups above a given valuation will be the same, but they will be conjugate, since the action of the Galois group is transitive. Given a choice of valuation w extending v, we have a corresponding extension L_w/K_v of complete fields.

Theorem 3.3. *The extension L_w/K_v is Galois, with Galois group $\mathrm{Gal}(L_w/K_v) = G_w$.*

Proof. See [14], Chapter II, §3. □

Given a Galois extension of local fields L/K, there is a corresponding extension of finite residue fields $\tilde{L}/\tilde{K}$, and the Galois group $\mathrm{Gal}(L/K)$ maps into $\mathrm{Gal}(\tilde{L}/\tilde{K})$. In the case discussed above, we define the inertia group $I_v \subset G_v$ to be the kernel of the reduction map on Galois groups. The inertia group will be trivial if and only if the extension is unramified (has ramification degree 1). Furthermore, the reduction map on Galois groups is surjective, so unramified extensions of local fields correspond to extensions of the residue field.

Next, we consider elliptic curves defined over local fields. let K be a local field and let E/K be an elliptic curve. By changing coordinates if necessary, we can find a Weierstrass equation for E all of whose coordinates are integral (that is, have nonnegative valuation); in this case, the discriminant, which is a polynomial in the coefficients, will also be integral. If the valuation of the discriminant is minimal across all Weierstrass equations with integral coefficients, we say that the Weierstrass equation is minimal. Given any point $[x : y : z] \in \mathbb{P}^2$, we can rescale coordinates so that all three are integral, and so that at least one has valuation zero. Since at least one reduced point will be nonzero, reducing these coordinates modulo the maximal ideal (π) gives a map:

$$E(K) \to \tilde{E}(\tilde{K})$$

Here, $\tilde{K}$ is the finite residue field of K, and $\tilde{E}$ is the curve whose equation is the reduced form of the equation for E. The discriminant of the equation corresponding to $\tilde{E}$ will be nonzero in $\tilde{K}$ if, and only if, the valuation of the original discriminant was zero.

Definition 3.4. Let E/K be an elliptic curve over a local field, and pick a minimal Weierstrass equation for E. We say that E *has good reduction* if the reduced curve $\tilde{E}$ is nonsingular. This is equivalent to requiring the discriminant of the minimal Weierstrass equation for E to have valuation 0, and does not depend on which minimal Weierstrass equation is chosen.

Usually, we will be interested in elliptic curves over a number field K. Given some E/K, we can interpret it as an elliptic curve over K_v for every valuation v on K. We say E has good reduction *at* v if the corresponding curve E/K_v has good reduction. For example, given an elliptic curve defined over $\mathbb{Q}$, we will often reduce its coefficients mod p, where p is a prime number.

By studying how the coordinates of Weierstrass equations transform, we can show that the discriminant only changes by 12th powers. This gives the following result:

Proposition 3.5 ([16], Remark VII.1.1)**.** *Let E be an elliptic curve over a local field K given by a Weierstrass equation. If the Weierstrass equation has integral coefficients, and the discriminant has valuation less than 12, then the Weierstrass equation is minimal.*

For example, if an elliptic curve $E/\mathbb{Q}$ has discriminant Δ, and Δ is not divisible by p^{12} for any prime p, then the corresponding Weierstrass equation is minimal at every prime p. In particular, E has good reduction at exactly the primes not dividing Δ.

The following result, which tells us how torsion points transform under reduction, is crucial in the proof of the Mordell-Weil Theorem:

Lemma 3.6. *Suppose E/K is an elliptic curve defined over a local field K, and suppose E has good reduction. Let p be the characteristic of its residue field. If m is coprime to p, then reduction gives an injective map:*

$$E(K)[m] \to \tilde{E}(\tilde{K})$$

In other words, the reduction map is injective on m-torsion, for m coprime to p.

Proof. See [16], Proposition VII.3.1. □

4. Galois Cohomology

Let E be an elliptic curve defined over the field K, and pick $\sigma \in \mathrm{Gal}(\bar{K}/K)$. The map σ acts on projective space by acting on its coordinates, sending $[x : y : z]$ to $[\sigma(x) : \sigma(y) : \sigma(z)]$. Let E^σ be defined by letting σ act on the coefficients of a Weierstrass equation for E. Then E^σ is an elliptic curve, and σ sends $E(\bar{K})$ to $E^\sigma(\bar{K})$. In particular, if E is defined over K, then σ induces a group automorphism of $E(\bar{K})$, although the map will *not* in general be a morphism of varieties.

The main technical apparatus in the proof of the Mordell-Weil Theorem is the analysis of $\mathrm{Gal}(\bar{K}/K)$-actions, especially this one. In this section, we briefly summarize the facts about Galois cohomology, which is a way to frame and simplify many facts about Galois actions. All facts quoted without proof in this section are from the appendix to [16]. For many more details, see the article [20].

Let K be a number field. The Galois group $\mathrm{Gal}(\bar{K}/K)$ has a natural topology as a profinite group, known as the *Krull topology*, which has the property that the subgroup $\mathrm{Gal}(\overline{K}/F)$, for F a finite Galois extension, form a neighborhood basis around the identity (for more details, see [17]). An Abelian group M endowed with an action of $\mathrm{Gal}(\bar{K}/K)$ is called a $\mathrm{Gal}(\bar{K}/K)$-module if the action is *continuous.* That is, the corresponding map $f\colon \mathrm{Gal}(\bar{K}/K) \times M \to M$ is a continuous map of topological spaces, where M is always taken with the discrete topology.

Proposition 4.1. *An action of $\mathrm{Gal}(\bar{K}/K)$ on an Abelian group M is continuous if, and only if, the stabilizer of any element $m \in M$ is an open subgroup of finite index.*

Proof. A standard fact about profinite groups is that the open subgroups are exactly the closed subgroups of finite index. Therefore, it's enough to prove an action is continuous if an only if all stabilizers are open.

First, suppose the action is continuous, and pick $m \in M$. The stabilizer of m is the set $\{\sigma \in \mathrm{Gal}(\bar{K}/K) : \sigma(m) = m\}$; since M has the discrete topology, this will be open in $\mathrm{Gal}(\bar{K}/K)$ if and only if $P = \{(\sigma, m) \in \mathrm{Gal}(\bar{K}/K) : \sigma(m) = m\}$ is open in $\mathrm{Gal}(\bar{K}/K) \times M$. However, $P = f^{-1}(m) \cap \{(x, m) : m \in M\}$, and the second set is open since M is discrete, so we conclude the stabilizer of m is open.

Conversely, suppose that for every $m \in M$, the stabilizer of m is open. We will show that the action map $f\colon \mathrm{Gal}(\bar{K}/K) \times M \to M$ is continuous. Since M has the discrete topology, it's enough to show that for every $m \in M$, $f^{-1}(\{m\})$ is open. We have:

$$f^{-1}(\{m\}) = (\sigma, n) : \sigma(n) = m\}$$

For any $a \in M$, set $B_a := f^{-1}(\{m\}) \cap (\mathrm{Gal}(\bar{K}/K) \times \{a\}) = \{(\sigma, a) : \sigma(a) = m\}$. It's enough to show that each B_a is open in the product, since $f^{-1}(\{m\})$ is the union of these sets. Since M is discrete, we find that B_a will be open in the product if and only if

$$C_a := \{\sigma : \sigma(a) = m\} \subset \mathrm{Gal}(\bar{K}/K)$$

is open in $\mathrm{Gal}(\bar{K}/K)$. If C_a is empty, then it's clearly open. Otherwise, pick some τ with $\tau(a) = m$. Then $\tau^{-1}(m) = a$. It follows that for any σ, $\sigma(a) = m$ if and only if $\tau(\sigma(a)) = a$. But then C_a is just the stabilizer of a, translated by τ. It follows from our assumption, along with the fact that $\mathrm{Gal}(\bar{K}/K)$ is a topological group, that C_a is open. This completes the proof. □

Given a $\mathrm{Gal}(\bar{K}/K)$-module M, the most basic question we can ask is what its fixed points are. Write:

$$M^{\mathrm{Gal}(\bar{K}/K)} = \{x \in M : \sigma(x) = x, \text{ for all } \sigma \in \mathrm{Gal}(\bar{K}/K)\}$$

We also write $M^{\mathrm{Gal}(\bar{K}/K)} = H^0(\mathrm{Gal}(\bar{K}/K), M)$, for reasons which will soon become clear.

Example 4.2. By the definition of $\mathrm{Gal}(\bar{K}/K)$, it acts on the Abelian group $\bar{K}$. This action is continuous: indeed, the stabilizer of any element x is just the kernel of the map $\mathrm{Gal}(\bar{K}/K) \to \mathrm{Gal}(F/K)$, where F is the Galois closure of $K(x)$. By Galois theory, this kernel is exactly $\mathrm{Gal}(\overline{K}/F)$, which is open by the definition of the Krull topology.

In this case, Galois theory says that the fixed points are exactly the elements of K. By almost the same argument, $H^0(\mathrm{Gal}(\bar{K}/K), E(\bar{K})) = E(K)$.

Suppose we have a short exact sequence of $\mathrm{Gal}(\bar{K}/K)$-modules:

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

This means that A, B and C are $\mathrm{Gal}(\bar{K}/K)$-modules, and that the maps f and g commute with the action of $\mathrm{Gal}(\bar{K}/K)$. It follows that there are corresponding maps $f\colon H^0(\mathrm{Gal}(\bar{K}/K), A) \to H^0(\mathrm{Gal}(\bar{K}/K), B)$ and $g\colon H^0(\mathrm{Gal}(\bar{K}/K), B) \to H^0(\mathrm{Gal}(\bar{K}/K), C)$, where we use the same letter by abuse of notation. The restriction of f will still be injective, and the kernel of the restriction of g will be the image of the restriction of f, so we can form a corresponding sequence:

$$0 \longrightarrow H^0(\mathrm{Gal}(\bar{K}/K), A) \xrightarrow{f} H^0(\mathrm{Gal}(\bar{K}/K), B) \xrightarrow{g} H^0(\mathrm{Gal}(\bar{K}/K), C)$$

If g were surjective, then H^0 would preserve exact sequences. Constructions which preserve exact sequences are immensely helpful, because having an exact sequence

often allows us to describe unfamiliar objects as quotients of ones we understand. Unfortunately, g need not be surjective. This suggests we should find a construction called H^1 which makes the following result true:

Theorem 4.3. *Let A, B, C be as above. Then there exists a natural map δ for which we have the following exact sequence:*

$$0 \longrightarrow H^0(\mathrm{Gal}(\bar{K}/K), A) \longrightarrow H^0(\mathrm{Gal}(\bar{K}/K), B) \xrightarrow{m} H^0(\mathrm{Gal}(\bar{K}/K), C) \xrightarrow{\delta} H^1(\mathrm{Gal}(\bar{K}/K), A) \longrightarrow H^1(\mathrm{Gal}(\bar{K}/K), B) \xrightarrow{m} H^1(\mathrm{Gal}(\bar{K}/K), C)$$

Luckily for us, such a construction exists, although its definition is not immediately intuitive. The following definition is from [16].

Definition 4.4. Let M be a $\mathrm{Gal}(\bar{K}/K)$-module. The group of continuous 1-cocycles from $\mathrm{Gal}(\bar{K}/K)$ to M, denoted by $Z^1_{cont}(\mathrm{Gal}(\bar{K}/K), M)$, is the group of maps $\xi\colon \mathrm{Gal}(\bar{K}/K) \to M$ which are continuous with respect to the discrete topology on M, and which satisfy the condition:

$$\xi(\sigma\tau) = (\xi(\tau))^\sigma + \xi(\tau)$$

For any $m \in M$, the map $\sigma \mapsto m^\sigma - m$ is a continuous cocycle. The group of coboundaries $B^1(\mathrm{Gal}(\bar{K}/K), M)$ is the subgroup of $Z^1_{cont}(\mathrm{Gal}(\bar{K}/K), M)$ consisting of elements of this form. The *first cohomology group* of M is the quotient:

$$H^1(\mathrm{Gal}(\bar{K}/K), M) = \frac{Z^1_{cont}(\mathrm{Gal}(\bar{K}/K), M)}{B^1(\mathrm{Gal}(\bar{K}/K), M)}$$

Note that when M is equipped with a *trivial* $\mathrm{Gal}(\bar{K}/K)$-action, the group $H^1(\mathrm{Gal}(\bar{K}/K), M)$ will just be $\mathrm{Hom}(\mathrm{Gal}(\bar{K}/K), M)$, the group of continuous homomorphisms. Indeed, in this case a cocycle is just a continuous homomorphism, and every coboundary is trivial. When working with (possibly infinite) Galois groups, the homomorphisms we consider will always be continuous homomorphisms.

With this definition, Theorem 4.3 becomes true. The required map δ is defined in the following way: pick $c \in H^0(\mathrm{Gal}(\bar{K}/K), C)$. Applying exactness at C, pick $b \in B$ such that $g(b) = c$. Then we define $\xi\colon \mathrm{Gal}(\bar{K}/K) \to B$ by:

$$\xi(\sigma) = b^\sigma - b$$

This function lands in the image of the mapping $A \to B$, which allows us to interpret ξ as an element of $H^1(\mathrm{Gal}(\bar{K}/K), A)$.

In later sections, we will occasionally omit the Galois group $\mathrm{Gal}(\bar{K}/K)$ to save space, when it is clear which group is acting.

There are two additional facts about Galois cohomology which we'll need. First, suppose we have a finite Galois extension L/K. Then if M is a $\mathrm{Gal}(\bar{K}/K)$-module, M is also a module for $\mathrm{Gal}(\bar{L}/L) = \mathrm{Gal}(\bar{K}/L)$, by restricting the Galois action to a subgroup. This gives a *restriction map*

$$\mathrm{Res}\colon H^1(\mathrm{Gal}(\bar{K}/K), M) \to H^1(\mathrm{Gal}(\bar{K}/L), M),$$

sending each cocycle to its restriction to the smaller group.

By Galois theory, $\mathrm{Gal}(L/K) \cong \mathrm{Gal}(\bar{K}/K)/\mathrm{Gal}(\bar{K}/L)$. Thus, the submodule $M^{\mathrm{Gal}(\bar{K}/L)}$ can be turned into a $\mathrm{Gal}(L/K)$-module.[2] Furthermore, if

$$\xi\colon \mathrm{Gal}(L/K) \to M^{\mathrm{Gal}(\bar{K}/L)}$$

is a cocycle, we can turn it into a cycle for $\mathrm{Gal}(\bar{K}/K)$ by composition:

$$\mathrm{Gal}(\bar{K}/K) \longrightarrow \mathrm{Gal}(L/K) \xrightarrow{\xi} M^{\mathrm{Gal}(\bar{K}/L)} \subset M$$

This is called an *inflation map.* We have the following:

Lemma 4.5 (Inflation-restriction sequence)**.** *With notation as above, there is an exact sequence*

$$0 \longrightarrow H^1(\mathrm{Gal}(L/K), M^{\mathrm{Gal}(\bar{K}/L)}) \xrightarrow{Inf} H^1(\mathrm{Gal}(\bar{K}/K), M) \xrightarrow{Res} H^1(\mathrm{Gal}(\bar{K}/L, M)$$

The only other specific fact we will need about Galois cohomology is this one:

Theorem 4.6 (Hilbert's Theorem 90)**.** *We have:*

$$H^1(\mathrm{Gal}(\bar{K}/K), \bar{K}^*) = 1$$

Proof. For a discussion of several proofs, see [19]. □

Example 4.7. As a first demonstration of the power of Galois cohomology, we give a quick proof of *Kummer theory*, which says that if a number field K contains all mth roots of unity, then any Galois extension L/K with Galois group $\mathbb{Z}/m\mathbb{Z}$ is of the form $L = K(a^{1/m})$ for some $a \in K$. We have the following exact sequence of $\mathrm{Gal}(\bar{K}/K)$-modules (using multiplicative notation):

$$1 \longrightarrow \mu_m \longrightarrow \bar{K}^* \xrightarrow{z \mapsto z^m} \bar{K}^* \longrightarrow 1$$

Taking Galois cohomology yields a long exact sequence, from which we pull the following:

$$K^* \xrightarrow{z \mapsto z^m} K^* \xrightarrow{\delta} H^1(\mathrm{Gal}(\bar{K}/K), \mu_m) \longrightarrow H^1(\mathrm{Gal}(\bar{K}/K), \bar{K}^*)$$

By Hilbert's Theorem 90, the final term is zero. This, along with exactness at the second term, implies that:

$$H^1(\mathrm{Gal}(\bar{K}/K), \mu_m) \cong K^*/(K^*)^m$$

Since we assumed that $\mu_m \subset K$, we have:

$$H^1(\mathrm{Gal}(\bar{K}/K), \mu_m) = \mathrm{Hom}(\mathrm{Gal}(\bar{K}/K), \mu_m) = \mathrm{Hom}(\mathrm{Gal}(\bar{K}/K), \mathbb{Z}/m\mathbb{Z})$$

In this case, the map δ is defined as follows: given $x \in K^*$, pick $y \in \bar{K}^*$ such that $y^n = x$ — in other words, take $y = x^{1/n}$. Then define ξ by $\xi(\sigma) = y^\sigma - y$. The kernel of this map will evidently be $\mathrm{Gal}(\bar{K}/K(x^{1/m}))$.

Suppose L/K is Galois with Galois group $\mathbb{Z}/m\mathbb{Z}$. Then there is a surjective homomorphism $f\colon \mathrm{Gal}(\bar{K}/K) \to \mathbb{Z}/m\mathbb{Z}$, whose kernel is $\mathrm{Gal}(\bar{K}/L)$. But by what we wrote above the kernel must be of the form $\mathrm{Gal}(\bar{K}/K(x^{1/m}))$ for some $x \in K^*$, with x defined up to mth powers. By Galois theory, we conclude $L = K(x^{1/m})$.

[2]The astute reader will complain that we only defined Galois cohomology for absolute Galois groups. However, a finite group is profinite with the discrete topology, and likewise all other definitions carry over without difficulty; this is because what we're doing is an (important) special case of the general study of group cohomology.

5. Selmer Groups and the Mordell-Weil Theorem

In this section, we outline the proof of the following result, mostly following chapters VIII and X of [16]:

Theorem 5.1 (Mordell-Weil)**.** *Suppose E/K is an elliptic curve defined over a number field K. Then its group of K-rational points $E(K)$ is a finitely-generated Abelian group.*

As we said in the introduction, this means that $E(K) \cong \mathbb{Z}^r \times F$, where r is a nonnegative integer and F is some finite group.

At the highest level, the proof of Mordell-Weil proceeds in two steps. First, we prove the *Weak Mordell-Weil Theorem*, which says that the group $E(K)/mE(K)$ is finite, for every $m > 1$. This condition certainly holds of any finitely generated Abelian group, but the converse doesn't hold; for instance $\mathbb{Q}/m\mathbb{Q}$ is trivial, even though $\mathbb{Q}$ isn't finitely-generated.

To bridge this gap, one introduces an object called a *height function*. Roughly speaking, this measures the complexity of a point with algebraic coordinates. Usually, we require that the height function transforms in predictable ways under summing of points, and that only finitely many points have height below a given bound; along with the Weak Mordell-Weil Theorem, these conditions force $E(K)$ to have a finite set of generators. However, we won't discuss height functions in any detail and instead refer the reader to [16].

To start, we consider the isogeny $[m] : E \to E$. Letting $E[m]$ denote all m-torsion points of E defined over $\bar{K}$, we have the following exact sequence:

$$0 \longrightarrow E[m] \longrightarrow E \xrightarrow{[m]} E \longrightarrow 0$$

In fact, this is an exact sequence of $\mathrm{Gal}(\bar{K}/K)$-modules, since all the maps are defined over K. Taking Galois cohomology gives us the following exact sequence:

$$0 \longrightarrow E(K)[m] \longrightarrow E(K) \xrightarrow{m} E(K) \xrightarrow{\delta} H^1(E[m]) \longrightarrow H^1(E) \xrightarrow{m} H^1(E)$$

Here, we leave the Galois group $\mathrm{Gal}(\bar{K}/K)$ out from our notation to save space. From the middle, we extract the following:

$$0 \longrightarrow E(K)/mE(K) \xrightarrow{\delta} H^1(E[m]) \longrightarrow H^1(E)[m] \longrightarrow 0$$

Here, $H^1(E)[m]$ denotes the m-torsion subgroup of $H^1(E)$. We call $H^1(E) = H^1(\mathrm{Gal}(\bar{K}/K), E)$ the *Weil-Châtelet group*, and write:

$$H^1(\mathrm{Gal}(\bar{K}/K), E) = WC(E/K)$$

Note that this depends on a choice of the field K.

Remark 5.2. The group $WC(E/K)$ has a fairly natural interpretation as a group of *homogeneous spaces*, which are curves isomorphic to E over $\bar{K}$, modulo isomorphisms defined over K. We won't discuss this: for more details see section X.3 of [16].

This sequence looks promising, since it injects $E(K)/mE(K)$ into another group we might hope to directly prove is finite. However, the middle term need not be finite: indeed, if $E[m] \subset E(K)$, then $H^1(\mathrm{Gal}(\bar{K}/K), E) = \mathrm{Hom}(\mathrm{Gal}(\bar{K}/K), E)$, which is not finite. To fix this, we repeat the same procedure with respect to local completions and combine the results. Let M_K denote a complete set of inequivalent absolute values on K: that is, M_K contains exactly one representative of each equivalence class of absolute values. The elements of M_K are also called *places*, and a place is called finite if it corresponds to a non-Archimedean absolute value.

Given any $v \in M_K$, we can extend v (non-uniquely) to $\bar{K}$, by Lemma 3.2. This allows us to define a completion $\bar{K}_v$, into which $\bar{K}$ naturally embeds. By abuse of notation, we let v denote the valuation on $\bar{K}$, chosen arbitrarily to extend the valuation on K. Then we have a decomposition group $G_v \subset \mathrm{Gal}(\bar{K}/K)$, which is the group of elements $\sigma \in \mathrm{Gal}(\bar{K}/K)$ that preserve the valuation. By Theorem 3.3, $G_v = \mathrm{Gal}(\bar{K}_v/K_v)$.

The group G_v, in its guise as $\mathrm{Gal}(\bar{K}_v/K_v)$, acts on $E(\bar{K}_v)$ in the obvious way. Indeed, since the previous argument used no facts about the field K, we can once again use Galois cohomology to construct an analogous sequence in the local case:

(5.3)
$$0 \longrightarrow E(K_v)/mE(K_v) \xrightarrow{\delta} H^1(G_v, E[m]) \longrightarrow H^1(G_v, E)[m] \longrightarrow 0$$

Note also that the last term is equal to $WC(E/K_v)[m]$. Combining all the local diagrams, we have the following commutative diagram:

(5.4)
$$\begin{array}{ccccccccc}
0 & \longrightarrow & E(K)/mE(K) & \xrightarrow{\delta} & H^1(\mathrm{Gal}(\bar{K}/K), E[m]) & \longrightarrow & WC(E/K)[m] & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \prod_{v \in M_K} E(K_v)/mE(K_v) & \xrightarrow{\delta} & \prod_{v \in M_K} H^1(G_v, E[m]) & \longrightarrow & \prod_{v \in M_K} WC(E/K_v) & \longrightarrow & 0
\end{array}$$

Here, the first vertical arrow comes from inclusion (modulo the appropriate quotient), and the other two come from restriction of a cocycle to the subgroup G_v. One can check directly from the definitions that this commutes.

By (5.4), the Weak Mordell-Weil Theorem exactly says that the kernel of the map $H^1(\mathrm{Gal}(\bar{K}/K), E[m]) \to WC(E/K)[m]$ is finite. We will in fact show slightly more.

Definition 5.5. The *m-Selmer group* of E/K is the kernel of the map
$$H^1(\mathrm{Gal}(\bar{K}/K), E[m]) \to \prod_{v \in M_K} WC(E/K_v),$$
defined in (5.4). By commutativity, either path defines the same map.

We also define the *Shafarevich-Tate group*, $Ш(E/K)$, to be the kernel of the map $WC(E/K) \to \prod WC(E/K_v)$.[3]

We will show the m-Selmer group is finite; as (5.4) makes clear, this is stronger than showing the kernel of the map $H^1(\mathrm{Gal}(\bar{K}/K), E[m]) \to WC(E/K)[m]$ is finite. The technical heart of this proof is the following. For its proof, which is mostly an exercise in algebraic number theory, see Proposition VIII.1.6 of [16].

[3]An open conjecture states that $Ш(E/K)$ is always finite.

Proposition 5.6. *Let K be a number field, let $S \subset M_K$ be a finite set of places containing all the infinite places, and pick an integer $m \geq 2$. Let L/K be the maximal abelian extension such that:*

(1) L has exponent m: that is, every element of $\mathrm{Gal}(L/K)$ has order dividing m.
(2) L/K is unramified outside of S.

Then L/K is finite. □

To see how Proposition 5.6 will help us, we will state a couple of additional definitions.

Definition 5.7. Let M be any set which admits an action by $\mathrm{Gal}(\bar{K}/K)$, and let v be a finite place. We say M is unramified at v if $I_v \subset \mathrm{Gal}(\bar{K}/K)$ acts trivially on M. Now, suppose M is an Abelian group. We say $\xi \in H^1(\mathrm{Gal}(\bar{K}/K), M)$ is unramified at v if the restriction of ξ to $H^1(I_v, M)$ is trivial (recall that ξ is a map from $\mathrm{Gal}(\bar{K}/K)$ to M).

Remark 5.8. We can motivate this definition in the following way: let L/K be a finite Galois extension of number fields, and define $M = \mathrm{Gal}(L/K)$. Since $\mathrm{Gal}(L/K)$ is a quotient of $\mathrm{Gal}(\bar{K}/K)$, it admits a natural action. The extension L/K will be unramified at the prime corresponding to v if and only if the inertia group $I_v \subset L/K$ is trivial. But this will hold if and only if the inertia subgroup *of the absolute Galois group* acts trivially on L.

Now, let $S \subset M_K$ be any finite set of places, assumed to contain all the infinite places. We define:

$$H^1_S(\mathrm{Gal}(\bar{K}/K), M) = \{\xi \in H^1(\mathrm{Gal}(\bar{K}/K), M) : \xi \text{ is unramified at all } v \notin S\}.$$

Remark 5.9. The group $H^1_S(\mathrm{Gal}(\bar{K}/K), M)$ is often *also* called a Selmer group, and its definition more or less corresponds to doing cohomology "subject to local restrictions." The key fact is that these restricted cohomology groups are always finite.

Lemma 5.10. *Let M be a finite abelian group with an action of $\mathrm{Gal}(\bar{K}/K)$, and let $S \subset M_K$ be a finite set of places containing all infinite places. Then $H^1_S(\mathrm{Gal}(\bar{K}/K), M)$ is always finite.*

Proof. We verified in Section 4 that since $\mathrm{Gal}(\bar{K}/K)$ acts continuously, the stabilizer of any given element is an open subgroup of finite index. Since M is finite, this means that the set of elements fixing all of M is an open subgroup of finite index, and in particular closed. By Galois theory, this subgroup will be of the form $\mathrm{Gal}(\bar{K}/K') = \mathrm{Gal}(\bar{K'}/K')$, for some finite extension K'/K. Applying inflation-restriction (Lemma 4.5), we have the following exact sequence:

$$0 \to H^1(\mathrm{Gal}(K'/K), M^{\mathrm{Gal}(\bar{K'}/K')}) \to H^1(\mathrm{Gal}(\bar{K}/K), M) \to H^1(\mathrm{Gal}(\bar{K'}/K'), M)$$

One can check that this sequence passes to Selmer groups, giving the following:

$$0 \to H^1_S(\mathrm{Gal}(K'/K), M^{\mathrm{Gal}(\bar{K'}/K')}) \to H^1_S(\mathrm{Gal}(\bar{K}/K), M) \to H^1_S(\mathrm{Gal}(\bar{K'}/K'), M)$$

The lefthand group is finite since both $\mathrm{Gal}(K'/K)$ and $M^{\mathrm{Gal}(\bar{K'}/K')}$ are finite. Therefore, to show the middle group is finite it's enough to show that the righthand group is. Replacing K with K', we may assume $\mathrm{Gal}(\bar{K}/K)$ acts trivially on M.

Suppose m has exponent M, where $m \geq 2$ (in other words, that $mx = 0$ for all $x \in M$). As in Proposition 5.6, define L/K to be the maximal Abelian extension of exponent m which is unramified outside of S. Then we know L/K is finite. From inflation-restriction, we have an inflation map:

$$\mathrm{Hom}_S(\mathrm{Gal}(L/K), M) \to \mathrm{Hom}_S(\mathrm{Gal}(\bar{K}/K), M)$$

In particular, we can see that pre-composition takes unramified cohomology classes to unramified cohomology classes. Recall that the first cohomology group corresponding to a trivial action is just a group of homomorphisms; thus it makes sense to consider Hom_S, even though we originally only defined the group H^1_S. To show the righthand group is finite, it's enough to show that the restriction map

$$\mathrm{Hom}_S(\mathrm{Gal}(\bar{K}/K), M) \to \mathrm{Hom}_S(\mathrm{Gal}(\bar{L}/L), M)$$

is trivial.

To prove this, consider any $f \in \mathrm{Hom}_S(\mathrm{Gal}(\bar{K}/K), M)$. The kernel of f will be an open subgroup of $\mathrm{Gal}(\bar{K}/K)$ of finite index, hence also closed, and thus will be of the form $\mathrm{Gal}(\bar{K}/F)$, where F/K is a finite Galois extension. The extension F must have exponent m: indeed, by Galois theory $\mathrm{Gal}(F/K) = \mathrm{Gal}(\bar{K}/K)/\mathrm{Gal}(\bar{K}/F)$. Given any $x \in \mathrm{Gal}(\bar{K}/K)$, $f(mx) = mf(x) = 0 \in M$, so $mx \in \ker(f) = \mathrm{Gal}(\bar{K}/F)$, and this proves F has exponent m.

Furthermore, the extension F is unramified outside of S. Indeed, by assumption, the map f is unramified outside of S, so for all $v \notin S$, $f(I_v) = 0$ (since all group actions considered here are trivial). But this means that for any such v, $I_v \subset \mathrm{Gal}(\bar{K}/F)$, so I_v is mapped to 0 in $\mathrm{Gal}(F/K)$, which exactly means that F is unramified at v.

Now, however, we know that $F \subset L$, since it's an extension of exponent m, unramified outside S. This means $\ker(f) = \mathrm{Gal}(\bar{K}/F) \supset \mathrm{Gal}(\bar{L}/L)$. But this means that f is in the kernel of the restriction map described above. Thus, the restriction map is trivial, and by our discussion above this completes the proof. □

Lemma 5.11. *With notation as above, the m-Selmer group $S^m(E/K)$ is finite.*

Proof. Recall that $S^m(E/K) \subset H^1(\mathrm{Gal}(\bar{K}/K), E[m])$. Let S be a finite set of places containing all infinite places, all places v such that $v(m) > 0$, and all places v such that E/K does not have good reduction at v. By the previous lemma, it's enough to show that any $\xi \in S^m(E/K)$ is unramified outside S.

Pick some such ξ, and pick any $v \notin S$. Let I_v be the inertia group for v, contained in the decomposition group G_v. By the definition of the Selmer group as a kernel, we know that ξ maps to 0 in $WC(E/K_v)$. By the exact sequence (5.3), this means the restriction of ξ to G_v is in the image of the connecting map $\delta\colon E(K_v)/mE(K_v)$. By the definition of the connecting map, this means for some $P \in E(\bar{K}_v)$, we have:

$$\xi(\sigma) = \{P^\sigma - P\}$$

Recall that I_v is the kernel of the reduction map:

$$\mathrm{Gal}(\bar{K}/K) \to \mathrm{Gal}(\tilde{\bar{K}}_v/\tilde{K}_v)$$

Thus, if $\sigma \in I_v$, then letting $\tilde{E}_v$ be the reduction of E with respect to v (which is an elliptic curve since E has good reduction at v), we know that σ acts trivially on $\tilde{E}_v$. In particular, we have:

$$\widetilde{P^\sigma - P} = \tilde{P}^\sigma - \tilde{P} = \tilde{O}.$$

Furthermore, since $\xi \in H^1(\mathrm{Gal}(\bar{K}/K), E[m])$, $\xi(\sigma) = P^\sigma - P \in E[m]$ for every σ. By definition of S, $v(m) = 0$, so in particular m is coprime to the characteristic of the residue field. Applying Lemma 3.6, we find that $E(K)[m]$ is mapped injectively into $\tilde{E}_v$ by the reduction map Thus, since the reduction of $P^\sigma - P$ is trivial, we must have $P^\sigma = P$ for any $\sigma \in I_v$, so ξ is trivial restricted to I_v. This proves that ξ is unramified outside of S, which completes the proof □

Now, we are ready to prove our main result for this section.

Proof of the Weak Mordell-Weil Theorem. It follows from (5.4) that $E(K)/mE(K)$ bijects with the kernel of the map $H^1(\mathrm{Gal}(\bar{K}/K), E[m]) \to WC(E/K)[m]$. However, the kernel of this map is clearly contained in the kernel of the map

$$H^1(\mathrm{Gal}(\bar{K}/K), E[m]) \to \prod_{v \in M_K} WC(E/K_v)[m],$$

since the second map factors through the first. This latter kernel is exactly the m-Selmer group, which we have verified is finite. □

The remainder of the proof of the Mordell-Weil Theorem is based on the idea of a height function, as we said above, and is laid out in detail in Chapter VIII of [16].

Remark 5.12. In fact, the full Mordell-Weil Theorem says more than we stated here: given any *Abelian variety* A over a number field K, the group $A(K)$ is finitely generated. The proof of this stronger result has the same basic structure: first prove the Weak Mordell-Weil Theorem, then introduce a notion of height to finish the argument. However, the details are significantly more complicated. For a proof, see Yuri Manin's appendix to [11].

6. Two Examples

In this section, we illustrate the proof of the Mordell-Weil Theorem with two specific examples. First, we apply techniques described in Chapter X of [16] to compute the rational points (the *Mordell-Weil group*) of a specific elliptic curve. Then, in order to illustrate some of the possible difficulties, we discuss a harder and more pathological example which will end up being important in our study of modular curves.

The computational strategy we will describe works with curves all of whose 2-torsion points are rational (recall that by Lemma 2.10, there are four 2-torsion points over an algebraic closure, including the identity). It can also be extended to curves which only have one 2-torsion point aside from the identity. However, as our second example will indicate, things can very quickly become difficult when neither assumption holds.

Here is an outline of what we do, omitting most of the proofs. Let E/K (where K for us will usually be $\mathbb{Q}$) be an elliptic curve defined by an equation

$$y^2 = x^3 + Ax^2 + Bx + C.$$

The first step is to compute the K-rational torsion on E. For simplicity, assume that $K = \mathbb{Q}$. By changing coordinates, we may always assume that the equation for E is integral. Then except for finitely many bad primes dividing the discriminant, the reduction $\tilde{E}$ of this curve mod p will be an elliptic curve over $\mathbb{F}_p$, and according to Lemma 3.6, whenever m is coprime to p this map will be injective on m-torsion.

By reducing modulo several different small primes and computing the resulting group $\tilde{E}(\mathbb{F}_p)$, we can easily bound the size of the torsion group, and then we only need to find generators for the torsion points whose existence we can't exclude.

For example, suppose we reduce a curve mod 2 and find that it has 5 points, and then reduce it mod 5 and find that it has 5 points. Then we know the m-torsion for m coprime to 2 is a subgroup of $\mathbb{Z}/5\mathbb{Z}$ by the first reduction, and we likewise know by the second reduction that the 2-torsion is trivial. It follows that the torsion subgroup of $E(\mathbb{Q})$ is either 0 or $\mathbb{Z}/5\mathbb{Z}$, and if the first case holds then we should be able to verify the torsion subgroup is trivial by reducing modulo additional primes.

Once we know the torsion subgroup F, the harder task is to determine the rank r of the curve, giving a decomposition $E(\mathbb{Q}) \cong \mathbb{Z}^r \times F$. To determine the rank, it's enough to determine the size of the group $E(\mathbb{Q})/2E(\mathbb{Q})$.[4] Indeed, once we know the group F, we know how large $F/2F$ is, and the group $E(\mathbb{Q})/2E(\mathbb{Q})$ will have size $2^r \cdot |F/2F|$.

Therefore, it's enough to compute the size of $E(\mathbb{Q})/2E(\mathbb{Q})$. Since the procedure is not much more complicated in general, we describe it over an arbitrary number field K. Furthermore, we assume that all four 2-torsion points on E are K-rational. We observe from its equation (given above) that E is symmetric over the x-axis. Since the origin is $O = [0 : 1 : 0]$, this means that for every $P = (x, y) \in E(\mathbb{C})$, we have $-P = (x, -y)$. Thus, the 2-torsion points, aside from the origin, are the points for which $y = 0$. These are exactly the roots of the righthand side of the defining equation, which tells us that it factors:

$$x^3 + Ax^2 + Bx + C = (x - e_1)(x - e_2)(x - e_3), e_1, e_2, e_3 \in K$$

We will let $S \subset M_K$ be a finite set of "bad places" (recall that M_K is a complete set of absolute values on K). In this case, S will contain the infinite places, the places corresponding to primes that divide 2, and all places corresponding to primes at which E has bad reduction. We define:

$$K(S, 2) = \{x \in K^*/(K^*)^2 : v(x) \equiv 0 \bmod 2, \text{ for all } v \notin S\}$$

By the fundamental theorem of Kummer Theory, we find that:

$$K(S, 2) \cong H^1_S(\mathrm{Gal}(\bar{K}/K), \{\pm 1\})$$

In particular, an element ξ of the cohomology group is unramified at a given $v \notin S$ if, and only if, the corresponding $x \in K^*/(K^*)^2$ satisfies $v(x) \equiv 0 \bmod 2$. This isomorphism tells us that the group $K(S, 2)$ is finite, by the results proved in the previous section.

Recall that the group $E[2]$ of 2-torsion points on an elliptic curve is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ as a group. Furthermore, since all 2-torsion points on E are assumed rational, the Galois action is trivial, so this isomorphism is also an isomorphism of trivial $\mathrm{Gal}(\bar{K}/K)$-modules. One checks once again that the conditions imposed for places $v \notin S$ line up appropriately, and then we have:

$$\begin{aligned} H^1_S(\mathrm{Gal}(\bar{K}/K), E[2]) &\cong H^1_S(\mathrm{Gal}(\bar{K}/K), \{\pm 1\}) \times H^1_S(\mathrm{Gal}(\bar{K}/K), \{\pm 1\}) \\ &\cong K(S, 2) \times K(S, 2) \end{aligned}$$

Recall that the Selmer group $S^2(E/K)$ is a subgroup of $H^1_S(\mathrm{Gal}(\bar{K}/K), E[2])$, and that $E(K)/2E(K)$ embeds into $S^2(E/K)$. By composition, this gives an injective

[4] However, finding specific generators for the group $E(\mathbb{Q})$ is harder, and this is not a problem we will deal with.

homomorphism $E(K)/2E(K) \hookrightarrow K(S,2) \times K(S,2)$. By tracing through the definitions, we can explicitly determine what this map is, and then computing the size of $E(K)/2E(K)$ reduces to computing the size of its image in $K(S,2) \times K(S,2)$. It turns out that determining whether a given element is in the image amounts to finding rational solutions to a pair of polynomial equations.

Theorem 6.1 ([16], Proposition X.1.4)**.** *Let E/K, S, and $K(S,2)$ be as above. Then there is an injective homomorphism*

$$E(K)/2E(K) \to K(S,2) \times K(S,2)$$

defined by

$$P = (x,y) \mapsto \begin{cases} (x-e_1, x-e_2) & \text{if } x \neq e_1, e_2, \\ \left(\frac{e_1-e_3}{e_1-e_2}, e_1-e_2\right) & \text{if } x = e_1, \\ \left(e_2-e_1, \frac{e_2-e_3}{e_2-e_1}\right) & \text{if } x = e_2, \\ (1,1) & \text{if } P = O \end{cases}$$

Suppose $(b_1,b_2) \in K(S,2) \times K(S,2)$ is not the image of $O, (e_1,0)$, or $(e_2,0)$. Then (b_1,b_2) is in the image of the mapping if and only if the equations:

$$b_1 z_1^2 - b_2 z_2^2 = e_2 - e_1$$
$$b_1 z_1^2 - b_1 b_2 z_3^2 = e_3 - e_1$$

have a solution $(z_1,z_2,z_3) \in K^* \times K^* \times K$.

If we get lucky, then given each pair (b_1,b_2), we can either find a point that maps to it, or show that the corresponding system of equations is insoluble over some completion of K . However, it can happen that a given system of equations is "locally" solvable in every completion of K, but not solvable in K itself. This is one reason why the procedure describes does not give a guaranteed algorithm to compute the Mordell-Weil group.

Example 6.2. We illustrate the procedure described above, by computing the Mordell-Weil group on the curve

$$E : y^2 = x^3 - 3x^2 + 2x = x(x-1)(x-2)$$

It is clear from the equation that E has all 2-torsion rational, and computing its discriminant we find that it has good reduction at all primes except 2. Reducing mod 3, we find that the curve contains the four points $O, (0,0), (0,1)$ and $(0,2)$. Therefore, the torsion subgroup of $E(\mathbb{Q})$ is either $(\mathbb{Z}/2\mathbb{Z})^2 \times \mathbb{Z}/3\mathbb{Z}$, or $(\mathbb{Z}/2\mathbb{Z})^2$. To exclude the first possibility, we reduce mod 5: the resulting curve has 8 points, so there cannot be any 3-torsion. Therefore, the torsion subgroup of $E(\mathbb{Q})$ is $(\mathbb{Z}/2\mathbb{Z})^2$.

Since E has good reduction away from 2, S will contain only the infinite places and the prime 2. This allows us to explicitly compute:

$$\mathbb{Q}(S,2) = \{\bar{1}, \bar{2}, \bar{-1}, \bar{-2}\}$$

Here, the bars denote equivalence classes. It remains only to check which elements of $K(S,2) \times K(S,2)$ are in the image. The points $(1,1), (2,1), (1,-1)$, and $(0,0)$ come from O, $(2,0)$, $(1,0)$ and $(0,0)$ respectively. By considering the first equation, we can eliminate all four possibilities where $b_1 < 0$, $b_2 > 0$: in this case both $b_1 z_1^2$ and $-b_2 z_2^2$ will be non-positive, and hence the equation has no solutions in $\mathbb{R}$. A similar argument eliminates the four cases where $b_1 < 0$, $b_2 > 0$, leaving only four cases to check.

In fact, we only need to check one more case, namely $(1, -2)$. This is because the map in question is a homomorphism, and hence its image is a subgroup: if any of the points $(1, 2), (2, 2)$ or $(2, -2)$ were in the image, then multiplying by other elements of the image would give that $(1, -2)$ is as well. So we are reduced to studying the equations:

$$z_1^2 + 2z_2^2 = 1, z_1^2 + 2z_3^2 = 2$$

We will show these equations have no solution in $\mathbb{Q}_2$. Indeed, from the first equation we find that $v(1) = 0 = v(z_1^2 + 2z^2) = \min(2v(z_1), 2v(z_2) + 1)$. The final equality follows because the two valuations are necessarily different — one is even, the other odd — in which case a basic fact about $\mathbb{Q}_p$ is that their sum will take the smaller valuation. Since 0 is even, we find $v(z_1) = 0$, $v(z_2) \geq 0$. However, the second equation gives

$$v(2) = 1 = \min(2v(z_1), 1 + 2v(z_3)) = \min(0, 1 + 2v(z_3)) = 0,$$

where the final equality follows from the same basic fact. This is a contradiction, so the equations have no 2-adic solutions.

It follows that the image of $E(\mathbb{Q})/2E(\mathbb{Q})$ has size 4, which means that the group $E(\mathbb{Q})/2E(\mathbb{Q})$ itself has size 4. We conclude that the rank is 0, which gives:

$$E(\mathbb{Q}) \cong (\mathbb{Z}/2\mathbb{Z})^2$$

In this case, computing the Mordell-Weil group proved not to be very hard. However, as we said, no general method is known. Here is an important example of a curve E for which the group $E(\mathbb{Q})$ is not so easy to compute.

Example 6.3. Consider the elliptic curve $E : y^2 + y = x^3 - x^2$ Using the procedure described above, we can easily compute the torsion subgroup of $E(\mathbb{Q})$, which turns out to be $\mathbb{Z}/5\mathbb{Z}$. However, with the methods described in this section it will be difficult to go much further, since none of this curve's 2-torsion points are rational.

The procedure described above is known as 2-descent. As we said above, a somewhat more complicated version, called descent by 2-isogeny, can be used when only one two-torsion point is rational. We can also generalize the procedure to $m \neq 2$, but for larger m the process becomes significantly more complicated and the equations become much harder to work with.

Alternatively, we could pass to a finite extension $K/\mathbb{Q}$ such that $E[2] \subset E(K)$, and directly apply the procedure applied above. However, this raises its own difficulties. In particular, it may be difficult to determine how much larger the group $E(K)$ is, compared to the group $E(\mathbb{Q})$.

With more work, one can overcome these issues; it turns out that this curve has rank 0, so the Mordell-Weil group is $\mathbb{Z}/5\mathbb{Z}$ (this curve is indexed as 11.a3 in the database [18]).

7. Modular Curves

We now change direction substantially, in order to discuss the following question:

Question 7.1. *Let E be an elliptic curve defined over $\mathbb{Q}$. What are the possible torsion subgroups (necessarily finite) of the group $E(\mathbb{Q})$?*

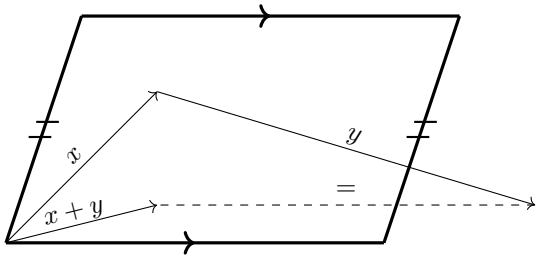
On the basis of the theory we've developed so far, this is a very difficult question to answer. As the examples developed in the previous section show, computing the torsion subgroup of a specific curve is not difficult, and by applying a little

ingenuity we can often determine the full group of rational points. But these results say nothing about which torsion subgroups could potentially occur across all elliptic curves.

The solution to this problem is called a *modular curve*, and can be thought of as a quotient of the upper half plane which parameterizes classes of elliptic curves along with additional structure. In order to define modular curves, we first discuss the analytic theory of elliptic curves over the complex numbers, skipping most proofs. For the rest of the paper, we assume basic background in complex analysis, including the definitions of holomorphic and meromorphic functions and the statement of Liouville's Theorem. However, analysis is not our focus and those willing to take things on faith should be able to follow along.

7.1. **Elliptic Curves as Riemann Surfaces.** Let $E/\mathbb{C}$ be any elliptic curve defined over the complex numbers (which includes those defined over $\mathbb{Q}$). Any nonsingular algebraic curve over $\mathbb{C}$ can be given the structure of a one-dimensional complex manifold, or Riemann surface. This is a consequence of the holomorphic implicit function theorem ([5], Chapter 0), and in any case fits well with our intuition that smooth curves are smooth.

In the case of elliptic curves, we can say much more: it turns out that every elliptic curve is isomorphic to a quotient of $\mathbb{C}$. First, some terminology. We say that an additive subgroup $\Lambda \subset \mathbb{C}$ is a *lattice* if Λ is generated by two complex numbers α, β which are not $\mathbb{R}$-linearly dependent. In this case, we observe that the quotient $\mathbb{C}/\Lambda$ is compact and topologically isomorphic to a torus. Furthermore, $\mathbb{C}/\Lambda$ is an Abelian group, since it's the quotient of $\mathbb{C}$ by a subgroup. The group law is given by addition of complex numbers, modulo the equivalence relation defined by the lattice. The fundamental parallelogram spanned by α and β surjects onto $\mathbb{C}/\Lambda$, giving rise to the following topology and group structure:



In complex analysis, a function which is periodic with respect to two $\mathbb{R}$-linearly independent periods α, β is called *doubly periodic*, or (for reasons that will soon become clear) *elliptic*. It is not hard to see that any nonconstant elliptic function cannot be everywhere holomorphic. Indeed, if it were, then the image of $\mathbb{C}$ would be the same as the image of the fundamental parallelogram, and hence compact. In particular, the function would be bounded, and thus constant by Liouville's Theorem. As a result, elliptic functions are only required to be meromorphic rather than holomorphic.

Given a lattice $\Lambda = \langle \alpha, \beta \rangle$, we could ask for a nice description of the elliptic functions which are periodic with respect to Λ. It turns out that such a description exists in terms of what are called Weierstrass $\wp$-functions.

Definition 7.2. Let $\Lambda \subset \mathbb{C}$ be a lattice. The *Weierstrass $\wp$-function relative to* Λ is defined by

$$\wp(z;\Lambda) = \frac{1}{z^2} + \sum_{\omega\in\Lambda,\omega\neq 0} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right).$$

Theorem 7.3. *Let $\Lambda \subset \mathbb{C}$ be a lattice. Then all elliptic functions are rational combinations of $\wp(z;\Lambda)$ and its derivative, $\wp'(z;\Lambda)$.*

Proof. See [16], Theorem VI.3.2. □

We will give no details on the theory of elliptic functions; for an introduction, see Chapter I of [15]. For our purposes, what makes $\wp$-functions interesting is that they give the Riemann surface $\mathbb{C}/\Lambda$ the structure of an elliptic curve, by acting as coordinate functions. Remarkably, this allows us to identify the group structure on $\mathbb{C}/\Lambda$ with the group structure on the corresponding elliptic curve.

Theorem 7.4. *Given a lattice Λ, there exists an elliptic curve $E/\mathbb{C}$, such that the map $\phi\colon \mathbb{C}/\Lambda \to \mathbb{P}^2$ defined by*

$$z \mapsto [\wp(z) : \wp'(z) : 1]$$

is an isomorphism both of Riemann surfaces and of groups onto $E(\mathbb{C})$.

Proof. See [16], Theorem VI.3.6. □

An even more remarkable fact is that we can go in the opposite direction. Combined, these two results will allow us to study complex elliptic curves by studying equivalence classes of lattices.

Theorem 7.5 (Uniformization theorem for elliptic curves)**.** *Let $E/\mathbb{C}$ be an elliptic curve. Then there exists a lattice $\Lambda \subset \mathbb{C}$, such that the map ϕ defined above identifies $\mathbb{C}/\Lambda$ with E.*

Proof. See [15], Chapter I, Corollary 4.3. □

This looks promising: lattices look easier to study than elliptic curves, and indeed they are. For instance, given lattices Λ, Λ', all the maps between them are given by linear functions.

Proposition 7.6. *Let Λ, Λ' be lattices. Then:*

- *(a) Every holomorphic map $\mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$ is of the form $\phi(z+\Lambda) = mz+b+\Lambda'$, where $m, b \in \mathbb{C}$ and $m\Lambda \subset \Lambda'$.*
- *(b) If ϕ is also a group homomorphism, then we may take $b = 0$. In particular, $\mathbb{C}/\Lambda \cong \mathbb{C}/\Lambda'$ as complex Lie groups if and only if $\Lambda = \gamma\Lambda'$ for some $\gamma \in \mathbb{C}^*$. In this case, we say the two lattices are* homothetic.

Proof. For the proof of (a), see Proposition 1.3.2 in [3]; the trick is to lift the map ϕ to the complex plane, and analyze it there. To prove (b), suppose ϕ is a group homomorphism. This means in particular that ϕ sends the equivalence class of 0 to the equivalence class of 0. But $\phi(0) = b$, so we must have $b \in \Lambda'$, and up to the quotient we obtain the same map by taking $b = 0$. Finally, if $\phi(z) = mz$, ϕ is clearly invertible if and only if $m \neq 0$, which shows the final statement. □

7.2. **The Action of $\mathbf{SL}_2(\mathbb{Z})$.** In this section we define a specific modular curve in detail. Everything we do will extend to the case of general modular curves, discussed in the next section.

Following [3], call a quotient $\mathbb{C}/\Lambda$ a *complex torus*. We now know that to study elliptic curves over $\mathbb{C}$, we only need to study complex tori, and to study complex tori, we only need to study lattices over $\mathbb{C}$ up to homothety. To do this, let's pick a basis $\langle \alpha, \beta \rangle$ for the lattice over $\mathbb{Z}$. Further, let's assume that the basis is positively oriented, or in other words that β/α has positive imaginary part (it should be clear that we can always pick a basis with this property by reversing the order if necessary).

Since we are free to rescale the lattice, we divide by α. This gives rise to a basis of the form $\langle 1, \tau \rangle$, and since $\langle \alpha, \beta \rangle$ was positively oriented, we find that $\mathrm{Im}(\tau) > 0$. We write $\mathbb{H}$ for the set of complex numbers $x + iy$ with $y > 0$.

Question 7.7. *Given two numbers $\tau, \tau' \in \mathbb{H}$ when are $\langle 1, \tau \rangle$ and $\langle 1, \tau' \rangle$ bases for the same lattice?*

Define $\mathrm{SL}_2(\mathbb{Z})$ to be the following group:

$$\mathrm{SL}_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{Z}, ad - bc = 1 \right\}$$

We will also use the notation $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$. For now, the number 1 only exists to create a sense of mystery; later, we'll define groups $\Gamma(N)$ for all N. $\mathrm{SL}_2(\mathbb{Z})$ has a natural action on the upper half plane, given by:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau = \frac{a\tau + b}{c\tau + d}$$

One can check ([15], Chapter I, Lemma 1.1) that this maps the upper half plane to itself.

If we want, we can view this as a special case of an action of $\mathrm{SL}_2(\mathbb{C})$ on the complex projective line, whose definition is the same as the one given above. The corresponding automorphisms are called *fractional linear transformations*; in general, the automorphisms of the projective line over a field k are given by the fractional linear tranformations corresponding to elements of $\mathrm{SL}_2(k)$. In our case, the action of the overall group $\mathrm{SL}_2(\mathbb{C})$ does not preserve the upper half plane, but the action of $\mathrm{SL}_2(\mathbb{R}) \subset \mathrm{SL}_2(\mathbb{C})$ does preserve it, so $\mathrm{SL}_2(\mathbb{Z})$ does as well. A version of this idea will come up later, when we discuss cusps.

For another perspective, observe that lattices are free $\mathbb{Z}$-modules (that is, Abelian groups) of rank 2. A group automorphism of $\mathbb{Z}^2$ exactly corresponds to an invertible 2×2 matrix with integer entries. In our specific setting, we require that the determinant be positive in order to preserve orientation, which suggests that we consider elements of $\mathrm{SL}_2(\mathbb{Z})$ (the only units in $\mathbb{Z}$ are ± 1). In order to make things look more familiar, we temporarily reverse the order of the basis to $\langle \tau, 1 \rangle$, and identify $\tau = (1, 0), 1 = (0, 1)$. Given $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$, we compute:

$$A^T((1,0)) = (a, b), A^T((0,1)) = (c, d)$$

Since we are free to rescale, this corresponds to the choice of oriented basis:

$$\left\{ 1, \frac{a\tau + b}{c\tau + d} \right\}.$$

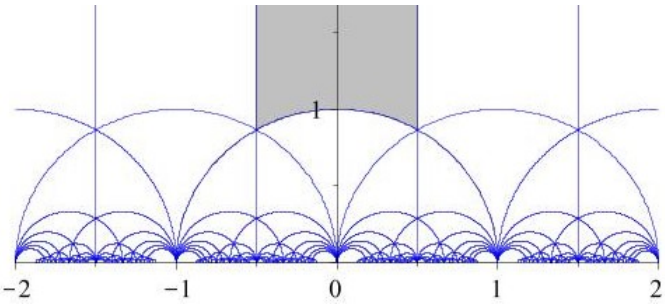


FIGURE 2. An illustration of the action of $\mathrm{SL}_2(\mathbb{Z})$ on the upper half plane. Image due to Wikimedia user Kilom691.

This suggests the action of $\mathrm{SL}_2(\mathbb{Z})$ is really just a change of basis in disguise. Writing Λ_τ for the lattice spanned by 1 and τ, we have the following:

Proposition 7.8. *Given $\tau, \tau' \in \mathbb{H}$, the lattices Λ_τ and $\Lambda_{\tau'}$ are homothetic if, and only if, there exists some $\phi \in SL_2(\mathbb{Z})$ such that $\phi(\tau) = \tau'$.*

Proof. In one direction, suppose that $\tau' = \phi(\tau)$ for some $\phi \in \mathrm{SL}_2(\mathbb{Z})$. By the discussion above, this means that $1, \tau'$ are a basis for the same lattice, up to rescaling. Conversely, suppose $\Lambda_{\tau'} = \gamma\Lambda_\tau$, where $\gamma \in \mathbb{C}^*$. This means that $\gamma, \gamma\tau$ form a basis (necessarily positively-oriented) for the lattice which is also spanned by $1, \tau'$. By our remark above about automorphisms of $\mathbb{Z}^2$, we know that $1 = c(\gamma\tau) + d\gamma$, $\tau' = a(\gamma\tau) + b\gamma$ for some $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. Dividing, we find as expected that $A(\tau) = \tau'$. □

It turns out, although we won't need to use this fact, that the group $\mathrm{SL}_2(\mathbb{Z})$ is generated by the matrices $\sigma = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tau = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Both of these matrices act in visually intuitive ways: the first translates $\tau \mapsto \tau + 1$, while the second maps $\tau \mapsto \frac{-1}{\tau}$, which amounts to reflection in the circle $\mathbb{S}^1$ followed by reflection over the y-axis.

We now know that the orbits of the $\Gamma(1)$-action on $\mathbb{H}$ correspond to the isomorphism classes of elliptic curves over $\mathbb{C}$. This makes it natural to define

$$Y(1) = \Gamma(1)\backslash\mathbb{H},$$

the quotient of the upper half plane by the action of $\Gamma(1)$.[5] The quotient $Y(1)$ has a natural quotient topology. In fact, although this is not obvious, more is true:

Fact 7.9. With an appropriate choice of charts, $Y(1)$ is a Riemann surface.

Let $\pi\colon \mathbb{H} \to Y(1)$ be the projection map. At most points $\pi(x) \in Y(1)$, the Riemann surface structure is easy to define, because no element of $\Gamma(1)$ fixes $\pi(x)$ aside from $\pm I$, both of which act trivially. Thus, we just take $\pi(U)$ for a sufficiently small neighborhood U of x, to obtain a local chart. However, there are two points for which this condition fails: i and the third root of unity ω_3. We call these points

[5]Since multiplying a matrix by ± 1 does not change its action on $\mathbb{H}$, some authors instead let the action be that of the quotient group $\mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$.

elliptic points; a bit more work is required to define the Riemann surface structure at these points, but it can be done.

Our ultimate hope is to apply tools from algebraic geometry in order to find rational points on modular curves such as $Y(1)$. The basis for this approach is the following theorem.

Theorem 7.10. *Every compact Riemann surface is an algebraic curve.*

Proof. See [5], among others. □

We can't immediately use Theorem 7.10, because the curve $Y(1)$ is not compact. To fix this, we will add finitely many rational points called *cusps*, which do not correspond to elliptic curves. This makes the resulting object worse at classifying curves, but far better behaved.

To see how to proceed, consider Figure 2. The shaded area, D, is called a *fundamental domain* for the group action, roughly because its translations under the group action form a tiling of $\mathbb{H}$. If D were compact, then its image under the quotient, which is $Y(1)$, would be as well. In fact, D fails to be compact; after making the appropriate identifications, we see that D becomes compact when the point $\infty = [1:0]$ on the Riemann sphere is added to it.

However, we cannot just add the point ∞, since we also need to consider the group action. Reflection in the unit circle sends the fundamental domain D to a wedge which tapers to the point 0, suggesting that this reflection should send ∞ to 0. In fact, for every $q \in \mathbb{Q}$, there is an element of the group which sends ∞ to q.

One way to see this is to recall that $\mathrm{SL}_2(\mathbb{Z})$ is really acting on the Riemann sphere by fractional linear transformations. In this case, the action is of the form

$$\infty \mapsto \frac{a \cdot 1 + b \cdot 0}{c \cdot 1 + d \cdot y} = \frac{a}{c}$$

where $ad - bc = 1$. By picking a fraction $q = \frac{a}{c}$ in lowest terms and using basic facts about principal ideal domains, we can always find an element of $\mathrm{SL}_2(\mathbb{Z})$ whose action is of the form given above.

This motivates the following initially bizarre-looking definition:

Definition 7.11. The modular curve $X(1)$ is the quotient of $\mathbb{H} \cup \{\infty\} \cup \mathbb{Q} \subset \mathbb{C}$, under the action of $\Gamma(1)$ by fractional linear transformations.

Points coming from $\mathbb{Q} \cup \{\infty\}$ are called *cusps*. In the case we're currently dealing with, the group action will actually collapse all the added points into a single orbit, as we just verified. However, we will soon deal with other modular curves which have multiple cusps.

Fact 7.12. $X(1)$ can be made into a compact Riemann surface, and thus is an algebraic curve.

Proof. See Chapter 2 of [3]. □

A final issue, which will become much more important in the next section, is the question of fields of definition. Suppose $E, E'/\mathbb{Q}$ are isomorphic over $\mathbb{C}$: then they are also isomorphic over $\bar{\mathbb{Q}}$. Indeed, by Fact 2.7, the curves will be isomorphic over an algebraically closed field if, and only if, they have the same j-invariant. Since both fields are algebraically closed, the result follows. However, the isomorphism *need not be defined over* $\mathbb{Q}$, and it need not preserve structure relative to $\mathbb{Q}$. For

instance, two elliptic curves defined over $\mathbb{Q}$ may have different groups of rational torsion points, despite being isomorphic over $\mathbb{C}$ (equivalently $\bar{\mathbb{Q}}$).

In our case, it turns out that if a point on $X(1)$ is rational then there is some $E/\mathbb{Q}$ contained in the corresponding isomorphism class. However, since multiple such curves may have different rational torsion structure, we want to be able to distinguish between them.

7.3. **Modular Curves and Mazur's Theorem.** In order to solve the problem just described, we now consider the action of a subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$.[6] The resulting modular curves will "remember more information" about the underlying curves, giving a finer classification. Three special classes of subgroups show up most often:

Definition 7.13. Pick $N \geq 1$. Then:

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod N \right\}$$

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \bmod N \right\}$$

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \bmod N \right\}$$

Note that for all N, $\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N)$. Likewise, note that

$$\Gamma(1) = \Gamma_1(1) = \Gamma_0(1) = \mathrm{SL}_2(\mathbb{Z}),$$

which explains our notation from the previous section.

Just as with $\Gamma(1)$, we can construct a corresponding modular curve. We call the corresponding curves $Y(N)$, $Y_1(N)$, and $Y_0(N)$ respectively. Likewise, we can compactify these curves, arriving at compact Riemann surfaces $X(N), X_1(N), X_0(N)$. However, the compactified curve may now have multiple cusps, depending on how $\{\infty\} \cup \mathbb{Q}$ decomposes into orbits. Knowing exactly how many cusps there are becomes critical for classifying torsion, because showing there is no rational n-torsion for some n amounts to showing that on some modular curve, the *only* rational points are the finitely many cusps which were added. Once we give our surfaces the structure of algebraic curves, the question of which cusps are rational becomes even more difficult, and depends on which algebraic curve model is used; for more details on this, see the discussion in [13].

It turns out that the orbits of the above subgroups correspond to information about torsion on curves.

Proposition 7.14. *Pick $N \geq 1$.*

- *(a) Points of $Y_0(N)$ correspond to pairs (E, G), where E is an elliptic curve and $G \subset E$ is a cyclic subgroup of order N. Two such pairs $(E, G), (E', G')$ are identified when there is an isomorphism of E onto E' taking G to G'.*
- *(b) Points of $Y_1(N)$ correspond to pairs (E, P), where E is an elliptic curve and $P \in E$ is a point of exact order N. Two such pairs $(E, P), (E', P')$ are identified when there is an isomorphism of E onto E' taking P to P'.*

[6] One usually requires the subgroup to be a *congruence subgroup*, or in other words to contain $\Gamma(N)$ for some N. However, all subgroups we consider will satisfy this condition.

(c) *Points of $Y(N)$ correspond to triples (E, P, Q), where E is an elliptic curve and P, Q are two points which generate group $E[N]$ of N-torsion points. Two triples $(E, P, Q), (E', P', Q')$ are identified when there is an isomorphism of E onto E' taking P to P' and Q to Q'.*

Now that we have added additional structure, issues around fields of definition become even more subtle. For instance, it turns out that rational points of $Y_0(N)$ need not correspond to actual rational torsion points: if a given cyclic subgroup $G \subset E$ is invariant under the Galois group, then G will be defined over $\mathbb{Q}$ even if none of its nontrivial points are, and the point corresponding to (E, G) will be rational. However, when studying the family of curves $X_1(N)$, we get lucky.

Theorem 7.15. *Suppose $N \geq 4$, and suppose $(E, P), (E', P')$ are pairs corresponding to rational points on $Y_1(N)$. The two pairs correspond to the same point if, and only if, there is an isomorphism $f\colon E \to E'$* defined over $\mathbb{Q}$ *which sends P to P'.*

Proof. This is a hard result. For a discussion, see [21], and for more details see Chapter 7 of [3]. □

Thus, determining whether there are rational N-torsion points on elliptic curves defined over $\mathbb{Q}$ is equivalent to determining whether the curve $X_1(N)$ has rational points which are not cusps.

We are now in a position to say something about the statement of Theorem 1.1. It's not very hard to compute the genus of the modular curve $X_1(N)$ for various N (for example, see Theorem 3.1.1 of [3]). Doing so, one finds that $X_1(N)$ has genus 0 for $1 \leq N \leq 10$ and $N = 12$, genus 1 for $N = 11, 14, 15$, and genus ≥ 2 for all other N. Observe that the genus 0 cases are exactly the possible torsion orders, according to the statement of Mazur's Theorem. This is not surprising: every modular curve has at least one rational cusp, and every genus 0 curve with a rational point is isomorphic over $\mathbb{Q}$ to the projective line, so in particular every such curve has infinitely many rational points.

In 1973, Andrew Ogg conjectured in [13] that these cases are the only possible cases — in other words, that there is a rational elliptic curve with a rational N-torsion point if, and only if, $X_1(N)$ has genus 0. Barry Mazur proved this conjecture, in two papers ([9],[7]) published in 1977 and 1978.

By a famous result known as Faltings's Theorem, any curve defined over $\mathbb{Q}$ of genus at least 2 has only finitely many rational points. This immediately gives the following step toward Mazur's theorem.

Theorem 7.16. *Suppose $N \geq 16$. Then there are only finitely many isomorphism classes of elliptic curves defined over $\mathbb{Q}$ with a rational N-torsion point.*

However, this does not provide a uniform bound, and thus does not prove the full theorem. In any case, Faltings's Theorem was proved after Mazur's Theorem, and its proof is at least as difficult.

Example 7.17. As a special case of Mazur's Theorem, consider the modular curve $X_1(11)$. This is a compact Riemann surface, hence an algebraic curve. In fact, it is of genus one and has a a rational point, so $X_1(11)$ is an elliptic curve over $\mathbb{Q}$. One possible Weierstrass equation for $X_1(11)$ is:

$$y^2 + y = x^3 - x^2$$

This is an equation we've seen before: as discussed in Example 6.3, its group of rational points is $\mathbb{Z}/5\mathbb{Z}$, of size 5. But one can check that the curve has 5 rational cusps. Therefore, $Y_1(11)$ contains no rational points, and we conclude that there are no rational 11-torsion points on elliptic curves defined over $\mathbb{Q}$.

For more details on this example from the perspective of modular curves, see the notes [21]. This fact can also be proven much more directly, by using the definition of addition on an elliptic curve and several changes of coordinates: see the classic paper [1], where the result was first proven, or the discussion in [22].

One can carry out a similar analysis on the other cases where $X_1(N)$ is an elliptic curve, and show that all rational points are cusps. Several other special cases were already known before Mazur's Theorem was proved: for example, see [8], which deals with the case $N = 13$, and [6], which bounds p^n-torsion for a fixed prime p.

Acknowledgments

Thanks to my mentor, Karl Schaefer, for answers to many, many questions, as well as fascinating discussions about elliptic curves, number theory, and many other adjacent areas of mathematics. Thanks to Matthew Emerton for a very enlightening conversation about modular forms, and to Brian Lawrence for helpful answers to some questions about line bundles. Thanks to Joseph Silverman, David Roberts, and Chris Wuthrich for helpful answers to questions on MathOverflow. Finally, thanks to Peter May for running a fantastic REU program.

References

[1] Billing, G. and K. Mahler. "On exceptional points on cubic curves". In: *J. London Math. Soc.* 15 (1940), pp. 32–43.

[2] Conrad, Keith. *Ostrowski for number fields*. URL: `https://kconrad.math.uconn.edu/blurbs/gradnumthy/ostrowskinumbfield.pdf` (visited on 08/19/2019).

[3] Diamond, Fred and Jerry Shurman. *A First Course in Modular Forms*. Springer-Verlag, New York, 2005.

[4] Fulton, William. *Algebraic Curves*. An introduction to algebraic geometry, Notes written with the collaboration of Richard Weiss, Reprint of 1969 original. 2008. URL: `http://www.math.lsa.umich.edu/~wfulton/CurveBook.pdf`.

[5] Griffiths, Phillip and Joseph Harris. *Principles of Algebraic Geometry*. Wiley Classics Library. Reprint of the 1978 original. John Wiley & Sons, Inc., New York, 1994.

[6] Manin, Ju. I. "The p-torsion of elliptic curves is uniformly bounded". In: *Izv. Akad. Nauk SSSR Ser. Mat.* 33 (1969), pp. 459–465.

[7] Mazur, B. "Rational isogenies of prime degree (with an appendix by D. Goldfeld)". In: *Invent. Math.* 44.2 (1978), pp. 129–162.

[8] Mazur, B. and J. Tate. "Points of order 13 on elliptic curves". In: *Invent. Math.* 22 (1973), pp. 41–49.

[9] Mazur, Barry. "Modular curves and the Eisenstein ideal". In: *Inst. Hautes Études Sci. Publ. Math.* 47 (1977). With an appendix by Mazur and M. Rapoport, 33–186 (1978). URL: `http://www.numdam.org/item?id=PMIHES_1977__47__33_0`.

[10] Milne, J. S. *Elliptic Curves*. BookSurge Publishers, Charleston, SC, 2006.

[11] Mumford, David. *Abelian Varieties*. With appendices by C. P. Ramanujam and Yuri Manin, Corrected reprint of the second (1974) edition. Tata Institute of Fundamental Research, 2008.

[12] Neukirch, Jürgen. *Algebraic number theory*. Translated from the 1992 German original and with a note by Norbert Schappacher, With a foreword by G. Harder. Springer-Verlag, Berlin, 1999.

[13] Ogg, Andrew P. "Rational points on certain elliptic modular curves". In: *Analytic number theory (Proc. Sympos. Pure Math., Vol XXIV, St. Louis Univ., St. Louis, Mo., 1972)*. 1973, pp. 221–231.

[14] Serre, Jean-Pierre. *Local fields*. Translated from the French by Marvin Jay Greenberg. Springer-Verlag, New York-Berlin, 1979.

[15] Silverman, Joseph H. *Advanced Topics in the Arithmetic of Elliptic Curves*. Springer-Verlag, New York, 1994.

[16] Silverman, Joseph H. *The Arithmetic of Elliptic Curves*. Second. Vol. 106. Graduate Texts in Mathematics. Springer, Dordrecht, 2009. URL: https://doi.org/10.1007/978-0-387-09494-6.

[17] Szamuely, Tamás. *Galois groups and fundamental groups*. Cambridge University Press, Cambridge, 2009.

[18] The LMFDB Collaboration. *The L-functions and Modular Forms Database*. http://www.lmfdb.org. Online; accessed 19 August 2019.

[19] Various. *Motivation for the proof of Hilbert's Theorem 90*. URL: https://mathoverflow.net/q/73077/140821 (visited on 08/20/2019).

[20] Washington, Lawrence C. "Galois cohomology". In: *Modular forms and Fermat's last theorem (Boston, MA, 1995)*. Springer, New York, 1997, pp. 101–120.

[21] Weston, Tom. *The modular curves $X_0(11)$ and $X_1(11)$*. 1999. URL: http://people.math.umass.edu/~weston/oldpapers/mc.ps (visited on 08/08/2019).

[22] Woodbury, Michael Carter. *Finite groups on elliptic curves*. URL: https://www.math.utah.edu/~woodbury/research/files/ellipticwriteup.pdf (visited on 08/21/2019).