# INTEGER FORMS AND ELLIPTIC CURVES

YUZU IDO, IAN RUOHONIEMI, AND MATTHEW STEVENS

Abstract. In this paper we develop the necessary background to understand Bhargava's 2015 result on the average rank of elliptic curves. We do this through expanding on integer forms, specifically quadratic and quartic forms, before going into depth on Galois cohomology and its role in elliptic curves.

Contents



## 1. Introduction

Few topics in algebra have as deep a historical establishment as that of integer forms. Quadratic forms, the degree two case of integer forms, have been studied since antiquity. Their theory was extensively developed by

*Date*: August 2019.

the work of Euler and Legendre, and then formalized and made rigorous in Gauss' *Disquisitiones*. However, progress then decreased in this branch of algebra, as other topics took the forefront.

A much more recent topic in algebra is elliptic curves, which has been in vogue since the early twentieth century. The main theorem in the basic theory of elliptic curves, the Mordell-Weil theorem, was not proven until 1929, and much remains unknown about the behavior of the elliptic curve group. One of the Millennium Prize questions, the Birch Swinnerton-Dyer Conjecture, deals with determining the rank of elliptic curves, one of the properties of the elliptic curve group. It is into this large open topic that integer forms have made their return. In his recent publications Bhargava has developed a connection between counting quartic forms and the average size of the Selmer group, and he has used this to prove novel and insightful results about the average rank of elliptic curves [1]. This paper shall lay the groundwork for understanding elliptic curves and quadratic forms well enough for a study of Bhargava's result.

This paper is generally organized into two large categories. As an introduction to integer forms, the first half is dedicated to the theory of quadratic forms. The second half shall focus on the theory of elliptic curves and its connection to quartic forms through the Selmer and Tate-Shafarevich groups. In the first half, the general properties of quadratic forms shall be enumerated. We will focus on when they represent zero (§2), what invariants they have (§3), and how many equivalence classes there are as the invariants vary, looking at binary quadratic forms specifically (§4). In the second half, after developing the basic theory of elliptic curves (§5), we will proceed to talk about some of the fundamental results in their group structure (§6). After this, at the expense of assuming some category theory we will develop the connection between quadratic forms and elliptic curves, by introducing Galois cohomology (§7) and other important geometric machinery (§8). We then apply Galois cohomology to the Tate-Shafaravich and Selmer groups (§9), and then finally show how counting quartic forms can give insight into the Selmer group, and therefore the average rank of elliptic curves (§10).

## 2. When Do Forms Represent Zero?

To begin our discussion of integer forms, we need to first define what they are.

**Definition 2.1.** *An **integer form** is a homogeneous polynomial with integer coefficients.*

For example, $x^2 + 3xy + 4y^2$ is an integer form. Since it is of degree two, we call it a quadratic form, and since it is in two variables, we know additionally that it is a binary quadratic form. Similarly, we call $x^2 + 4yz$ a ternary quadratic form, and $x^3 + 2x^2y - 4y^3$ a binary cubic form. In this section, we will focus on when these forms **represent zero**, that is, when the form $f$ has a nontrivial set of variables $x_1, x_2, ..., x_n$ such that

$f(x_1, ..., x_n) = 0$. We shall first consider a general statement for quadratic forms, Hasse-Minkowski Theorem, before narrowing our focus to an earlier result of Gauss.

2.1. **Serre's Proof of Hasse-Minkowski.** A common concern of both algebra and number theory is finding when equations have rational solutions. Finding real solutions is often comparatively easy, and finding solutions modulo a prime is also not too difficult. Fortunately, there is a way to obtain a rational solution given a real solution and a solution modulo every prime. However, before we can formalize that, we need to introduce the concept of the $p$-adics.

Recall the definition of the standard Euclidean metric on $\mathbb{Q}$: for $x, y \in \mathbb{Q}$, $d(x, y) = |x - y|$, where $|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0 \end{cases}$. When we complete $\mathbb{Q}$ with this metric we get the real numbers. However, suppose we use a different norm. For a given prime $p$, every $a \in \mathbb{Q}$ can be represented uniquely as $p^n \frac{b}{c}$, where $n \in \mathbb{Z}$ and neither $b$ nor $c$ are divisible by $p$. Then we can define $|a|_p = p^{-n}$, with $|0| = 0$ by fiat. Intuitively, a number is $p$-adically small if it is highly divisible by $p$. Showing that the $p$-adic norm satisfies all the properties of a norm is left as an exercise for the reader.

**Definition 2.2.** *The field $\mathbb{Q}_p$, called the field of $p$-**adics**, is defined as the completion of $\mathbb{Q}$ using the metric $d(x, y) = |x - y|_p$.*

This field is actually a PID, and as such has very nice properties, including unique factorization. There is also a very close connection between the $p$-adics and values modulo $p^n$.

**Lemma 2.3** (Hensel)**.** *Let $f(x)$ be a polynomial with integer coefficients, and let $p$ be some prime. Given some $\alpha_0$ such that $f(\alpha_0) \equiv 0 \mod p$ and $f'(\alpha_0) \not\equiv 0 \mod p$, the infinite sequence $\alpha_{n+1} = \alpha_n - \frac{f(\alpha_n)}{f'(\alpha_n)}$. Converges to a value $\alpha$ that is a root of $f$. Each $\alpha_n$ satisfies $f(\alpha_n) \equiv 0 \mod p^n$.*

This method parallels Newton's method for finding a root to a polynomial, except that it functions in the $p$-adics. We also see that as we increase $n$, our solutions modulo $p^n$ get closer and closer to the $p$-adic solution.

We shall now state the main result of this section. Recall a polynomial represents zero if it has a nontrivial root. We then have the following theorem:

**Theorem 2.4** (Hasse-Minkowski)**.** *A quadratic form $f$ represents 0 in $\mathbb{Q}$ if and only if it represents 0 in $\mathbb{Q}_p$ for all primes $p$ and it represents 0 in $\mathbb{R}$.*

We will use a proof of Serre [16] to illustrate this result. Before that, though, we need to introduce one additional bit of machinery.

**Definition 2.5.** *Let $a, b$ be nonzero elements of $\mathbb{Q}$. Then the value*

$$(a, b)_\nu := \begin{cases} +1 & \text{if } z^2 = ax^2 + by^2 \text{has a solution in } (\mathbb{Q}_p)^3 \\ -1 & \text{if } z^2 = ax^2 + by^2 \text{has no solution in } (\mathbb{Q}_p)^3 \end{cases}$$

*is called the* ***Hilbert symbol*** *of $a$ and $b$.*

The Hilbert symbol satisfies some very nice properties. First, $(a,b)_\nu = (b,a)_\nu$. Second, $(a,bc)_\nu = (a,b)_\nu \cdot (a,c)_\nu$. Third, as $\nu$ varies, $(a,b)_\nu$ is almost always $+1$. Finally, $\prod_p (a,b)_\nu = 1$. Proofs of these properties can be found in chapter III of [16]. These properties will play an important role in proving Hasse-Minkowski for the quaternary case. With these definitions, we are now ready to prove Theorem 2.3.

*Proof.* Let $f$ be written in diagonal form, then normalized so that the first coefficient is 1. Sufficiency in this formulation is clear, so we will focus on necessity, by considering binary, ternary, and quaternary forms separately, and then using induction to prove all higher forms.

*Binary:* Let $f = X^2 - aY^2$. For there to be a real solution, we see that $a$ must be positive. We write $a$ in terms of its prime decomposition, as follows: $a = \prod_p p^{\nu_p(a)}$. We consider the solution in each $\mathbb{Q}_p$; we find $\left(\frac{X}{Y}\right)^2 = a$, which since each $\mathbb{Q}_p$ is a UFD requires that $\nu_p(a)$ is even. Hence $a$ is a square, so there is a solution in $\mathbb{Q}$, $(\sqrt{a}, 1)$.

*Ternary:* We construct $f = X^2 - aY^2 - bZ^2$, where $a, b$ are squarefree. We now do induction on $|a| + |b|$, assuming $|b|/ge|a|$. If $|b| + |a| = 2$, we see $f = X^2 \pm aY^2 \pm bZ^2$, which has a clear rational solution every time it has a real solution. For the higher cases, we see $|b| \geq 2$, so we let $b = \pm p_1 \cdot ... \cdot p_k$ where the $p_k$ are distinct primes. For a given $p_i$ we consider the solution in $\mathbb{Q}_{p_k}$ modulo $p_i$, and find $X^2 - aY^2 \equiv 0 \mod p_i$. If $Y \equiv 0 \mod p_i$, then $X \equiv 0 \mod p_i$, and furthermore since $b$ is squarefree we see $Z \equiv 0 \mod p_i$. We can thus divide $(X, Y, Z)$ by $p_i$ until we find an $(X', Y', Z')$ where $Y \not\equiv 0 \mod p_i$. Then $a \equiv \left(\frac{X'}{Y'}\right)^2 \mod p_i$, so $a$ is a square mod each $p_i$, and hence $a$ is a square mod $b$. Thus there exist $t, b'$ such that $t^2 = a + bb'$. We choose $t$ such that $|t| \leq |b|/2$. Since $bb' = t^2 - a$, we see $bb'$ is a norm of $k(\sqrt{a})$ for $k = \mathbb{Q}$ or $k = \mathbb{Q}_|$ for all primes $p$. Thus $f$ represents 0 in $k$ if and only if $f' = X^2 - aY^2 - b'Z^2$ represents 0. However, based on our choice of $t$ we see that $b' < b$, so the inductive argument is complete.

*Quaternary:* We write $f = aX^2 + bY^2 - (cZ^2 + dW^2)$. For each $\nu$ we find an $x_\nu$ that is represented by both $aX^2 + bY^2$ and $cZ^2 + dW^2$. This implies that $(x_\nu, -ab)_\nu = (a,b)_\nu$ and $(x_\nu, -cd)_\nu = (c,d)_\nu$ for all primes $\nu$. By theorem 4 in chapter III of [16], this is exactly sufficient to generate a single value $x \in \mathbb{Q}^*$ that satisfies $(x, -ab)_\nu = (a,b)_\nu$ and $(x, -cd)_\nu = (c,d)_\nu$ for all primes $\nu$. Therefore $aX^2 + bY^2 + xT^2$ represents zero at all local places, so by the tertiary case it represents 0 in $\mathbb{Q}$. Following the same logic with $cZ^2 + bW^2 + xT^2$ we find values of $X, Y, Z, W \in \mathbb{Q}$ such that $x$ is represented by $aX^2 + bY^2$ and $cZ^2 + dW^2$, which shows that $f$ represents 0.

*General Case:* We write $f = h - g$, where $h = k_1X_1^2 + k_2X_2^2$ and $g = k_3X_3^2 + ... + k_nX_n^2$. We consider the set of all primes that divide the coefficients in $g$, along with 2 and $\infty$. This is a finite set; for each $\nu$ in this set we can

find an $a_\nu$ that is represented by both $h$ and $g$ in $\mathbb{Q}_\nu^*$. Since the squares of $\mathbb{Q}_\nu^*$ form an open set, by approximation theorem this implies the existence of $x_1, x_2 \in \mathbb{Q}$ such that with $a = h(x_1, x_2)$, we have $a/a_\nu$ is a square in $\mathbb{Q}_\nu^*$ for all $\nu$. We consider the form $f_1 = aZ^2 - g$, which represents 0 in all $\mathbb{Q}_\nu$ and is of a lesser order. Thus by the inductive hypothesis $g$ represents $a$ in $\mathbb{Q}$, so $f$ represents 0 in $\mathbb{Q}$.

□

2.2. **Gauss's Article 294 Algorithm.** Before the development of the $p$-adics and long before Hasse and Minkowski, the question of when polynomials have integer (and for homogeneous polynomials, equivalently rational) solutions was a topic of deep interest. Gauss was able to prove the following theorem.

**Theorem 2.6** (Gauss,[7])**.** *Let $a, b, c$ be relatively prime, nonzero, squarefree integers. Then the quadratic form*

$$f = ax^2 + by^2 + cz^2 \tag{2.1}$$

*will represent zero if and only if $-bc, -ac, -ab$ are squares modulo $a, b, c$ respectively, and not all of $a, b, c$ have the same sign.*

Incredibly, this is a special case of the ternary case of Hasse-Minkowski! If $a, b, c$ do not all have the same sign, then a real solution exists, and the results about modular squares are implied by solutions over all $p$-adics for $p|abc$. To see this, we consider $-bc$ and $a$. By Chinese Remainder Theorem we can combine the $p$-adic solutions for each $p|a$ to get a single solution to the equation modulo $a$. Let this solution be $x_0, y_0, z_0$. We then see that

$$by_0^2 + cz_0^2 \equiv 0 \mod a \Rightarrow -bc \equiv (c\frac{z_0}{y_0})^2 \mod a \tag{2.2}$$

which is exactly what we set out to show. We can construct the results for $-ac$ and $-ab$ identically.

2.2.1. *Proof of Theorem 2.5.* We will first illustrate the necessity of the conditions given. Suppose $p, q, r$ is some nontrivial solution to $f = 0$. We can assume without loss of generality that the three values are relatively prime integers. Furthermore, we observe that they are relatively prime to each other. If $p$ and $q$ had a common divisor $\mu$, then $cr^2 \equiv 0 \mod \mu^2$, which since $c$ is squarefree implies $r \equiv 0 \mod \mu$, a contradiction. We find that $-ap^2$ is represented by $by^2 + cz^2$, so by article 154 in Disquisitiones the determinant $-bc$ is a square modulo $ap^2$, and thus a square modulo $a$. Identical logic shows that $-ac$ is a square modulo $b$ and $-ab$ is a square modulo $c$.

Proof of sufficiency will be split into two parts: first, we will show that $f$ can be transformed to a form where all the terms have coefficients divisible by $abc$; second, we will construct a solution to this quadratic form, and thus construct a solution to $f$.

2.2.2. *Finding an Equivalent Form.* In order to find a form which satisfies the above condition, we need to first find $A, B, C$ such that $A$ is relatively prime to $b$ and $c$, $B$ is relatively prime to $a$ and $c$, and $C$ is relatively prime to $a$ and $b$. We do this by finding values that satisfy the following conditions:

$$A \equiv c \mod bA \equiv \sqrt{-ab} \mod c \tag{2.3}$$

$$B \equiv a \mod cB \equiv \sqrt{-bc} \mod a \tag{2.4}$$

$$C \equiv b \mod aC \equiv \sqrt{-ac} \mod b \tag{2.5}$$

This is guaranteed to be possible by the fact that $a, b, c$ are relatively prime. We notice

$$A^2a + B^2b + C^2c \equiv -b^2c + c^2b \equiv 0 \mod a \tag{2.6}$$

So $A^2a + B^2b + C^2c$ is divisible by $a$ and likewise by $b$ and $c$. If $A, B, C$ happen to have a common divisor, we factor out by it, and since it will be relatively prime to $abc$, it will not affect the above property.

We now construct three values $\alpha, \beta, \gamma$ such that $\alpha Aa + \beta Bb + \gamma Cc = 1$. Following this, we use the method in article 279 of Disquisitiones to find $\alpha', \alpha'', \beta', \beta'', \gamma', \gamma''$ satisfying

$$\beta'\gamma'' - \gamma'\beta'' = Aa \tag{2.7}$$

$$\gamma'\alpha'' - \alpha'\gamma'' = Bb \tag{2.8}$$

$$\alpha'\beta'' - \beta'\alpha'' = Cc \tag{2.9}$$

Let $f$ have matrix representation $M$, and let $d = -abc$ be the determinant of $f$. If we then define

$$N = \begin{bmatrix} d\alpha & \alpha' & \alpha'' \\ d\beta & \beta' & \beta'' \\ d\gamma & \gamma' & \gamma'' \end{bmatrix} \tag{2.10}$$

We see that every term of $N^TMN$ is divisible by $d$. By the properties of the entries, we see that $\det(N) = d$, hence the determinant of $N^TMN$ is $d^3$.

2.2.3. *Finding a Rational Point.* Article 277 demonstrates a way to send any indeterminate form of discriminant 1 to the form $x^2 + 2yz$. We can thus use this method to find matrix $L$ such that

$$L^T \frac{N^TMN}{d} L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{2.11}$$

We thus find $L^TN^TMNL$ represents $d^2x^2 + 2dyz$, so if we plug in $(0, 1, 0)$ or $(0, 0, 1)$ into this form, we get a solution. This is equivalent to saying that the last two columns of $NL$ are representations of 0 in $f$.□

2.2.4. *The Implementation.* We successfully wrote an implementation of the algorithm Gauss gave in article 294 in Python, the code for which can be found at github.com/AgentChicken/Article294. Since the method Gauss gave was entirely constructive, writing the implementation was not too challenging, although since it followed Gauss' method it failed to include some more modern optimizations. As a proof of concept, though, it successfully illustrated the practicality and innovation of Gauss' work.

## 3. Invariants of Binary Forms

Define $V_D$ as the vector space of binary forms of degree $D$ with integer coefficients. For instance,

$$V_2 = \{ax^2 + bxy + cy^2 \mid a, b, c \in \mathbb{Z}\}$$

Then, for a field $K$, the linear groups $SL_D(K)$ and $GL_D(K)$ have a group action on $V_D$ such that, for $M$ in the linear group and $f(x, y)$ a binary form,

$$M \cdot f(x, y) = f(Mx, My)$$

Different aspects of the literature choose different linear groups to define the notion of equivalence of forms according to convenience and context; the sections in this paper that refer to equivalence will make it explicit which linear group is under consideration. Two forms $f$ and $g$ are equivalent if there exists $M$ in the chosen linear group such that $M \cdot f = g$. We can interpret $M$ as an invertible change of variables. This relationship divides binary forms into a number of equivalence classes.

Invariants are quantities that are unchanged through such changes of variables; notably, they are the same regardless of which linear group is used to define equivalence. Formally, we define $\mathcal{I}_D$ as the set of all polynomials $P$ in the coefficients of a degree $D$ form $f \in V_D$ such that $P(M \cdot f) = P(f)$ for all $f \in V_D$. $\mathcal{I}_D$ admits a ring structure with polynomial addition and multiplication.

**Theorem 3.1.** *In the binary quadratic case, for $f(x, y) = ax^2 + bxy + cy^2$, the discriminant $\Delta = b^2 - 4ac$ is an invariant.*

**Proof** We prove this in the case of equivalence under $SL_2(\mathbb{Z})$; the other cases are morally similar.

A polynomial is invariant under $SL_2(\mathbb{Z})$ if and only if it is invariant under the actions of the generators of $SL_2(\mathbb{Z})$, $S := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $T := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. $S$ transforms $ax^2+bxy+cy^2$ into $cx^2-bxy+ay^2$ and $T$ transforms $ax^2+bxy+cy^2$ into $ax^2 + (2a + b)xy + (a + b + c)y^2$; thus, it must be that if $f(a, b, c)$ is an invariant,

$$f(a, b, c) = f(c, -b, a)$$

$$f(a, b, c) = f(a, 2a + b, a + b + c)$$

Indeed, $b^2 - 4ac = (-b)^2 - 4ca = (2a + b)^2 - 4a(a + b + c)$. □

**Remark 3.2.** *In fact, Hilbert shows using differential equations that the discriminant is the unique invariant in the sense that* $\mathcal{I}_2 = \mathbb{Q}[\Delta] \subset \mathbb{Q}[a, b, c]$ *[8].*

In the binary cubic case, for $f(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$, with $a, b, c, d \in \mathbb{Z}$, the discriminant $\Delta = b^2c^2 + 18abcd - 4ac^3 - 4b^3d - 27a^2d^2$ is again the unique invariant.

The binary quartic case is more complicated because the ring of invariants is generated by two independent invariants. For $f(x, y) = ax^4 + bx^3y + cx^2y^2 + dxy^3 + ey^4$, they are:

$$I = 12ae - 3bd + c^2$$

$$J = 72ace + 9bcd - 27ad^2 - 27eb^2 - 2c^3$$

That is,

$$\mathcal{I}_4 = \mathbb{Q}[I, J] \subset \mathbb{Q}[a, b, c, d, e]$$

[1]

## 4. Binary Forms with Bounded Discriminant

In this section, we discuss Dirichlet's class number formula. This gives an explicit count of the number of classes of binary quadratic forms of a given discriminant up to $SL_2\mathbb{Z}$ equivalence. The formula is a striking result in elementary number theory that relates the class number, an algebraic object, to an $L$-function, an analytic one.

**Remark 4.1.** *The analogous statement to Dirichlet's formula in the context of elliptic curves is the Birch and Swinnerton-Dyer conjecture.*

We follow the exposition in Harold Davenport's *Multiplicative Number Theory*, Chapter 6, "Dirichlet's Class Number Formula" [2]. As does Davenport, we quote a few results regarding quadratic forms. The terms "binary quadratic form," "quadratic form," and "form" will be used interchangeably in this section.

Given a binary quadratic form $Q(x, y) = ax^2 + bxy + cy^2$, its discriminant $d$ is fundamental if and only if $d \equiv 1 \mod 4$ and is square-free, or $d = 4m$ where $m \equiv 2$ or $3 \mod 4$ and $m$ is square-free. Hereafter, the term "discriminant" implies "fundamental discriminant" unless otherwise specified.

Two quadratic forms $ax^2 + bxy + cx^2$ and $a'(x')^2 + b'x'y' + c'(y')^2$ of a given fundamental discriminant $d$ are equivalent up to the $SL_2(\mathbb{Z})$ action if there exists a unimodular substitution

$$x = \alpha x' + \beta y'$$

$$y = \gamma x' + \delta y'$$

with $\alpha, \beta, \gamma, \delta \in \mathbb{Z}$ such that $\alpha\delta - \beta\gamma = 1$. The number of equivalence classes under this relation for a given discriminant $d$ can be shown to be finite.

A form $Q(x, y) = ax^2 + bxy + cy^2$ with discriminant $d < 0$ is definite, meaning that $Q(x, y)$ always has the same sign for $(x, y) \neq (0, 0)$. Given a

positive definite $Q(x,y)$, the form $-Q(x,y)$ is negative definite; thus, half of forms with discriminant $d<0$ are positive definite and half are negative definite, and it suffices to consider the positive definite ones (with $a>0$).

A form $Q(x,y)=ax^2+bxy+cy^2$ with discriminant $d>0$ is indefinite. It can be shown to be equivalent to some form with $a>0$, which will be the representative of its class.

$h(d)$ denotes the number of classes of forms up to $SL_2(\mathbb{Z})$ equivalence with discriminant $d$; we assume the forms to be positive definite if $d<0$. $h(d)\in\mathbb{Z}_{>0}$ since the principal form of discriminant $d$,

$$\begin{cases} x^2-\frac{1}{4}dy^2 & \text{if } d\equiv 0 \mod 4 \\ x^2+xy-\frac{1}{4}(d-1)y^2 & \text{if } d\equiv 1 \mod 4 \end{cases}$$

always exists.

**Definition 4.2.** *The **Kronecker symbol** $\left(\frac{a}{n}\right)$ for integer $a,n$ is defined as follows:*

*For odd $p_i$, $\left(\frac{a}{p_i}\right)$ is the Legendre symbol*

$$\left(\frac{a}{p_i}\right)=\begin{cases} 0 & \text{if } a\equiv 0 \mod p_i \\ 1 & \text{if } a\not\equiv 0 \mod p_i \text{ and } a \text{ is a quadratic residue mod } p_i \\ -1 & \text{if } a \text{ is a quadratic non-residue mod } p_i \end{cases}$$

*For $n=up_1^{e_1}\cdots p_k^{e_k}\neq 0$, where $u=\pm 1$ and $p_i$ prime,*

$$\left(\frac{a}{n}\right)=\left(\frac{a}{u}\right)\prod_{i=1}^{k}\left(\frac{a}{p_i}\right)^{e_i}$$

$$\left(\frac{a}{0}\right)=\begin{cases} 1 & \text{if } a=\pm 1 \\ 0 & \text{if otherwise} \end{cases}$$

$$\left(\frac{a}{1}\right)=1$$

$$\left(\frac{a}{-1}\right)=\begin{cases} 1 & \text{if } a\geq 0 \\ -1 & \text{if otherwise} \end{cases}$$

$$\left(\frac{a}{2}\right)=\begin{cases} 0 & \text{if } 2\mid a \\ 1 & \text{if } a\equiv\pm 1 \mod 8 \\ -1 & \text{if } a\equiv\pm 3 \mod 8 \end{cases}$$

Dirichlet's class number formula (1839) gives the class number $h(d)$. The first step in Dirichlet's proof describes $h(d)$ in terms of

$$L(1,\chi)=\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{d}{n}\right)$$

The second step, whose proof will be omitted, gives $L(1,\chi)$ as a finite sum.

**Remark 4.3.** *$L(1,\chi)$ is one of the first examples of an L-function, which are central to the study of number theory. Such functions allow application of analytic methods to study arithmetic objects and also associate different objects such as elliptic curves and modular forms to each other. They are strongly tied to current research areas such as the Riemann Hypothesis.*

**Theorem 4.4.** ***(Dirichlet's Class Number Formula)***

$$h(d) = \frac{w|d|^{1/2}}{2\pi} L(1,\chi) \text{if } d < 0 \text{ where } w = \begin{cases} 2 & \text{if } d < -4 \\ 4 & \text{if } d = -4 \\ 6 & \text{if } d = -3 \end{cases}$$

$$h(d) = \frac{d^{1/2}}{\log \epsilon} L(1,\chi) \text{if } d > 0$$

$$L(1,\chi) = -\frac{\pi}{|d|^{3/2}} \sum_{m=1}^{|d|} m \left(\frac{d}{m}\right) \text{if } d < 0$$

$$L(1,\chi) = -\frac{1}{d^{1/2}} \sum_{m=1}^{d} \left(\frac{d}{m}\right) \log \sin \frac{m\pi}{d} \text{if } d > 0$$

**Proof of Dirichlet's class number formula (Step 1)**

We will only discuss the case when $d < 0$; for a treatment of $d > 0$, refer to Davenport.

To set up the proof, we first consider automorphs of forms, or the unimodular substitutions that bring a form to itself, the number of which we denote by $w$ when $d < 0$. There are always two obvious automorphs:

$$x = x', y = y'$$

$$x = -x', y = -y'$$

If $d < -4$, these are the only automorphs, so $w = 2$.

If $d = -4$, there is only 1 class, represented by the principal form $x^2 + y^2$, which also has the automorph

$$x = y', y = -x'$$

and its negative, so $w = 4$.

If $d = -3$, there is also only 1 class, represented by the principal form $x^2 + xy + y^2$, which also has the automorphs

$$x = -y', y = x' + y'$$

$$x = x' + y', y = -x'$$

and their negatives, so $w = 6$.

Next, we will consider the total number of representations of $n \in \mathbb{Z}_{>0}$ by forms in a representative set of given discriminant $d$. When $d < 0$, forms are positive definite, so $n$ can be represented in only a finite number of ways; we denote this number by $R(n)$.

We state without proof the following important result in the classical theory of quadratic forms, developed by Lagrange and Gauss:

**Theorem 4.5.** *For $n > 0$ and $(n, d) = 1$*

$$R(n) = w \sum_{m|n} \left(\frac{d}{m}\right)$$

*where*

$$w = \begin{cases} 2 & \text{if } d < -4 \\ 4 & \text{if } d = -4 \\ 6 & \text{if } d = -3 \\ 1 & \text{if } d > 0 \end{cases}$$

In Step 1, Dirichlet essentially uses this expression to find the average value of $R(n)$ as $n$ varies; it suffices to consider $n$ such that $(n, d) = 1$.

$$\begin{aligned} w^{-1} \sum_{\substack{n=1 \\ (n,d)=1}}^{N} R(n) &= \sum_{\substack{n=1 \\ (n,d)=1}}^{N} \sum_{m|n} \left(\frac{d}{m}\right) \\ &= \sum_{\substack{n=1 \\ (n,d)=1}}^{N} \sum_{m_1 m_2 = n} \left(\frac{d}{m_1}\right) \\ &= \sum_{\substack{m_1 m_2 \le N \\ (m_1 m_2, d)=1}} \left(\frac{d}{m_1}\right) \\ &= \sum_{m_1 \le \sqrt{N}} \left(\frac{d}{m_1}\right) \sum_{\substack{m_2 \le N/m_1 \\ (m_2,d)=1}} 1 + \sum_{\substack{m_2 < \sqrt{N} \\ (m_2,d)=1}} \sum_{\sqrt{N} < m_1 \le N/m_2} \left(\frac{d}{m_1}\right) \end{aligned}$$

The last equality follows from the fact that the first double sum considers $(m_1, m_2)$ such that $m_1 \le \sqrt{N}$ and the second considers $(m_1, m_2)$ such that $m_1 > \sqrt{N}$.

$$\sum_{\substack{m_2 \le N/m_1 \\ (m_2,d)=1}} 1 = \frac{N}{m_1} \frac{\varphi(|d|)}{|d|} + O[\varphi(|d|)]$$

Intuitively, there are $\frac{N}{m_1}$ possible candidates for $m_2$ and $\frac{\varphi(|d|)}{|d|}$ is the approximate proportion of those that are coprime to d. Thus,

$$\begin{aligned} \sum_{m_1 \le \sqrt{N}} \left(\frac{d}{m_1}\right) \sum_{\substack{m_2 \le N/m_1 \\ (m_2,d)=1}} 1 &= \sum_{m_1 \le \sqrt{N}} \left(\frac{d}{m_1}\right) \left(\frac{N}{m_1} \frac{\varphi(|d|)}{|d|} + O[\varphi(|d|)]\right) \\ &= N \frac{\varphi(|d|)}{|d|} \sum_{m_1 \le \sqrt{N}} \left(\frac{d}{m_1}\right) \frac{1}{m_1} + \sum_{m_1 \le \sqrt{N}} \left(\frac{d}{m_1}\right) O[\varphi(|d|)] \\ &= N \frac{\varphi(|d|)}{|d|} \sum_{m_1 \le \sqrt{N}} \left(\frac{d}{m_1}\right) \frac{1}{m_1} + O(\sqrt{N}) \end{aligned}$$

for fixed $d$ and arbitrarily large $N$.

$\sum_{\sqrt{N}<m_1\leq N/m_2}\left(\frac{d}{m_1}\right)$ is bounded, or $O(1)$, because $\left(\frac{d}{m_1}\right)$ is a non-principal character to the modulus $|d|$, meaning it produces some $+1$ and $-1$ that cancel out in the sum. Thus

$$\sum_{\substack{m_2<\sqrt{N}\\(m_2,d)=1}}\sum_{\sqrt{N}<m_1\leq N/m_2}\left(\frac{d}{m_1}\right)=O(\sqrt{N})$$

$$w^{-1}\sum_{\substack{n=1\\(n,d)=1}}^{N}R(n)=N\frac{\varphi(|d|)}{|d|}\sum_{m\leq\sqrt{N}}\left(\frac{d}{m}\right)\frac{1}{m}+O(\sqrt{N})$$

Thus,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{\substack{n=1\\(n,d)=1}}^{N}R(n)=w\frac{\varphi(|d|)}{|d|}\sum_{m=1}^{\infty}\left(\frac{d}{m}\right)\frac{1}{m}$$

$$(\star)\qquad =w\frac{\varphi(|d|)}{|d|}L(1,\chi)$$

Noting that $\frac{\varphi(|d|)}{|d|}$ is the proportion of $n$ such that $(n,d)=1$, this states that the average value of $R(n)$ as $n$ varies is $wL(1,\chi)$.

Now we find a different expression for the average value of $R(n)$ by using the definition of $R(n)$, the total number of representations of $n$ by the set of representative forms of discriminant $d$.

Define $R(n,f)$ as the number of representations of $n$ by a particular form $f$ with discriminant $d$.

$$R(n)=\sum_{f}R(n,f)$$

where $f$ runs over a representative set of forms with discriminant $d$, meaning there are $h(d)$ such $f$. When $d<0$,

$$\sum_{\substack{n=1\\(n,d)=1}}^{N}R(n,f)=|S|$$

where

$$S=\{(x,y)\in\mathbb{Z}^2\mid 0<ax^2+bxy+cy^2\leq N\text{ and }(ax^2+bxy+cy^2,d)=1\}$$

The second condition restricts $(x,y)$ to certain pairs $(x_0,y_0)$ of residue classes to the modulus $|d|$; it can be shown that there are $|d|\varphi(|d|)$ such pairs. Thus

$$S=\{(x,y)\in\mathbb{Z}^2\mid 0<ax^2+bxy+cy^2\leq N\text{ and }x\equiv x_0,y\equiv y_0\mod|d|\}$$

The first condition restricts $(x, y)$ to an ellipse centered at $(0, 0)$ which expands uniformly as $N \to \infty$. The area of the ellipse is

$$\frac{2\pi}{\sqrt{4ac - b^2}} N = \frac{2\pi}{|d|^{1/2}} N$$

and the number of lattice points within it is asymptotic to

$$\frac{1}{|d|^2} \frac{2\pi}{|d|^{1/2}} N$$

as $N \to \infty$. Accounting for the $|d|\varphi(|d|)$ possible $(x_0, y_0)$,

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \sum_{\substack{n=1 \\ (n,d)=1}} R(n, f) &= \frac{1}{|d|^2} \frac{2\pi}{|d|^{1/2}} |d|\varphi(|d|) \\ &= \frac{\varphi(|d|)}{|d|} \frac{2\pi}{|d|^{1/2}} \end{aligned}$$

Finally, we examine this equation along with $\star$.

$$\begin{aligned} h(d) &= \sum_f 1 \\ &= \frac{\sum_f \sum_{\substack{n=1 \\ (n,d)=1}} R(n, f)}{\sum_{\substack{n=1 \\ (n,d)=1}} R(n, f)} \\ &= \frac{\sum_{\substack{n=1 \\ (n,d)=1}} R(n)}{\sum_{\substack{n=1 \\ (n,d)=1}} R(n, f)} \\ &= \frac{\lim_{N\to\infty} \frac{1}{N} \sum_{\substack{n=1 \\ (n,d)=1}} R(n)}{\lim_{N\to\infty} \frac{1}{N} \sum_{\substack{n=1 \\ (n,d)=1}} R(n, f)} \\ &= \frac{w \frac{\varphi(|d|)}{|d|} L(1, \chi)}{\frac{\varphi(|d|)}{|d|} \frac{2\pi}{|d|^{1/2}}} \\ &= \frac{w|d|^{1/2}}{2\pi} L(1, \chi) \text{ for } d < 0 \quad \square \end{aligned}$$

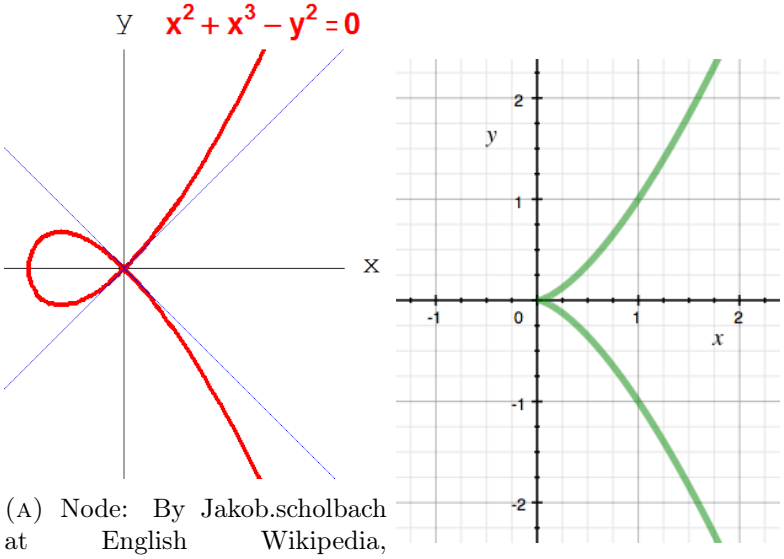
## 5. An Overview of Elliptic Curves

This section is a brief introduction to fundamental concepts in the study of elliptic curves. For interested readers, Silverman's *The Arithmetic of Elliptic Curves* gives a more detailed treatment of the subject [17].

**Definition 5.1.** *An **elliptic curve** defined over $\mathbb{Q}$, $E/\mathbb{Q}$, is a curve*

$$E = \{(x, y) \mid y^2 = f(x)\} \cup \{O\}$$

*with $f(x)$ a cubic with $\mathbb{Q}$ coefficients and distinct roots in $\mathbb{C}$, and $O$ the base point at infinity (to be explained later).*




(A) Node: By Jakob.scholbach at English Wikipedia, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=52468603

(B) Cusp: Public Domain, https://commons.wikimedia.org/w/index.php?curid=2267324

FIGURE 5.1. Node and cusp

With a suitable change of variables, an elliptic curve defined over $\mathbb{Q}$ can be written in the Weierstrass form

$$E : y^2 = x^3 + Ax + B \text{ with } A, B \in \mathbb{Q}$$

**Definition 5.2.** *Given the Weierstrass form, the* ***discriminant*** *is the quantity*

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

*associated to the elliptic curve.*

$f(x) = x^3 + Ax + B$ has distinct roots and $y^2 = f(x)$ is non-singular, if and only if $\Delta \neq 0$; else it has a node or a cusp (Figure 5.1) where the tangent is undefined and is not an elliptic curve.

**Definition 5.3.** *The* ***j-invariant*** *is the quantity*

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

Two elliptic curves are isomorphic over $\overline{\mathbb{Q}}$ if and only if they have the same $j$-invariant. That is, if two elliptic curves have the same $j$-invariant, we can do a change of variables involving coefficients in $\overline{\mathbb{Q}}$ which turns one into the other.

**Definition 5.4.** *The* ***set of rational points*** $E(\mathbb{Q})$ *is the set of points on the elliptic curve $E$ with $x$ and $y$ coordinates in $\mathbb{Q}$.*

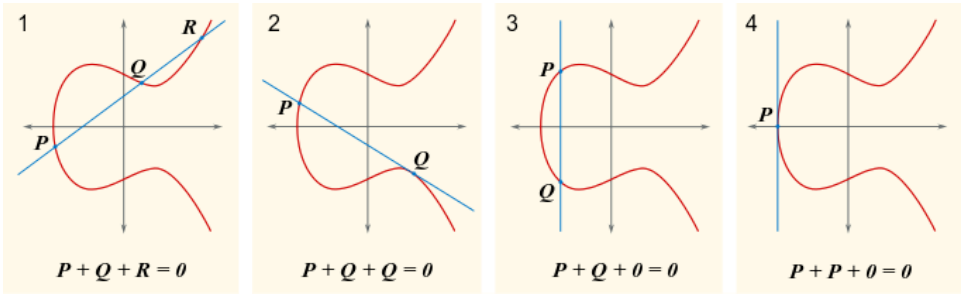



FIGURE 5.2. By SuperManu - Own work based on Image:ECClines.png by en:User:Chas zzz brown, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=2970559

$E(\mathbb{Q})$ has an abelian group structure based on the addition law depicted in the diagrams, where $O$ is the identity element (Figure 5.2). Since the curve is symmetric about the $x$-axis, the inverse of a point is its reflection across the $x$-axis. $mE(\mathbb{Q})$ is the subgroup of $E(\mathbb{Q})$ consisting of points that are $m$-multiples of points in $E(\mathbb{Q})$. For instance, in the second subfigure, $P + Q + Q = O$. This implies that $2Q = -P$ so, assuming $Q \in E(\mathbb{Q})$, $-P \in 2E(\mathbb{Q})$.

## 6. Finite Generation of Elliptic Curve Groups

The structure of the elliptic curve group has been a focus point within the study of elliptic curves. Central to this theory is the Mordell-Weil Theorem, which was proven in 1929 and shows that the elliptic curve group for an elliptic curve defined over any number field is finitely generated: specifically, that it takes the form $\mathbb{Z}^r \oplus T$ for some finite abelian group $T$, called the torsion group and $r$ is called the rank. Much of the behavior of torsion and rank remains unknown. Recent results on rank shall be addressed in sections 9 and 10 of this paper; the rest of this section will focus on an illustration of the Mordell-Weil Theorem for $\mathbb{Q}$, then discuss some of the more recent results in torsion.

### 6.1. **Mordell's Theorem.**

**Theorem 6.1.** ***(Mordell-Weil Theorem)*** *For an elliptic curve $E/\mathbb{Q}$, the group $E(\mathbb{Q})$ is finitely generated.*

The proof of this fact relies on two main parts: the Weak Mordell-Weil Theorem and the Descent Theorem.

#### 6.1.1. *Weak Mordell-Weil Theorem.*

**Theorem 6.2.** ***(Weak Mordell-Weil Theorem)*** *Let $m \geq 2$ be an integer. Then $E(\mathbb{Q})/mE(\mathbb{Q})$ is a finite group.*

To give an idea of how this result is proven we'll first write our elliptic curve in the form $y^2 = x(x-1)(x-\lambda)$, which is called Legendre form. Define the function $\Phi\colon E(\mathbb{Q}) \to \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \oplus \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ which satisfies $\Phi(x,y) = ([x],[x-1])$ for $x \neq 1, 0$ and maps 1 and 0 to $([1],[1])$. Notice that this is a group homomorphism, with kernel $2E(\mathbb{Q})$ (this can be verified with simple calculations). On $E/2E(\mathbb{Q})$, $\Phi$ is injective.

**Proposition 6.3.** *The image of $\Phi$ is finite.*

**Sketch**: Let $(x,y) \in E(\mathbb{Q})$ such that $x = m\cdot e^2, x-1 = n\cdot f^2$ for squarefree integers $m, n$ and rational $e, f$. Then the image of $(x,y)$ under $\Phi$ is $([m],[n])$. We observe $\frac{y}{ef}^2 = mn(x-\lambda)$. The only way the righthand side can be a square is if all the prime factors of $m$ divide $\lambda$ and all the prime factors of $n$ divide $\lambda - 1$. There are only a finite number of prime factors for $\lambda$ and $\lambda - 1$ and since $m$, $n$ are square-free, this makes there only a finite number of possibilities for $m, n$, so the image is finite.

Since $\Phi$ is injective, the finiteness of the image implies that $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite.

6.1.2. *Descent Theorem.*

**Definition 6.4.** *For a group $G$ we call any function $h : G \to \mathbb{R}$ where $h(g)$ represents the complexity of $g$ a* ***height function****.*

*Example* 6.5. The naive height function on the rationals can be described as follows: let $a = \frac{p}{q}$ where $p$ and $q$ are relatively prime. Then $H(a) = \max\{|p|, |q|\}$. We can refine this further to make it function multiplicatively by defining $h(a) = \log(H(a))$. This latter $h$ can also be used on $E(\mathbb{Q})$, acting on the $x$-coordinate of a given point.

*Example* 6.6. The height of an elliptic curve can also be defined. Let $E$ be in Weierstrass form, that is, it can be written as $y^2 = x^3 + Ax + B$. We then say $h(E) = \max\{|4A^2|, |27B^3|\}$.

The descent theorem takes an abelian group, like $E(\mathbb{Q})$, that admits a height function and shows from the height function that the group is finitely generated. For that to hold true, though, the height function must satisfy a few requirements.

**Theorem 6.7.** ***(Descent Theorem)*** *Suppose $A$ is an abelian group admitting a height function $h\colon A \to \mathbb{R}$ satisfying the following properties:*

(1) *Let $Q \in A$. $\exists C_1 \in \mathbb{R}$ dependent on $A$ and $Q$ such that*

$$h(P+Q) \leq 2h(P) + C_1 \ \forall P \in A \tag{6.1}$$

(2) *There exist integer $m \geq 2$ and constant $C_2$ such that*

$$h(mP) \geq m^2 h(P) - C_2 \ \forall P \in A \tag{6.2}$$

(3) *For every constant $C_3$ the set*

$$\{P \in A : h(P) \leq C_3\} \tag{6.3}$$

*is finite.*

*Additionally, suppose for the $m$ in (2.) $A/mA$ is finite. Then $A$ is finitely generated.*

**Proof of Theorem 7.4.** We choose $Q_1, Q_2, Q_3, ..., Q_r \in A$ as representatives for the separate cosets defined by $A/mA$, with $r$ being the number of cosets. We choose any $P \in A$. We construct $P_1$ such that $P = mP_1 + Q_{i_1}$ and $P_k$ such that $P_{k-1} = mP_k + Q_{i_k}$, with $i_k \in \{1, 2, ..., r\}$ based on $P$ (Aside: we see this construction because $P_{k-1} - Q_{i_k} \in mA$ if we choose the right $i_k$). We define the $P_k$ up to $k = n$ for some sufficiently large $n$. For any index $j$ we have

$$\begin{array}{rcl} h(P_j) & \leq & \frac{1}{m^2}(h(mP_j) + C_2) \\ & \leq & \frac{1}{m^2}(2h(P_{j-1}) + C_1' + C_2) \end{array} \tag{6.4}$$

Where $C_1'$ is defined by using property 1. on $-Q_1, -Q_2, ..., -Q_r$ and taking the largest resulting $C_1$, and $C_2$ is defined as it is in property 2.. We apply this inequality repeatedly and get

$$h(P_n) < \left(\frac{2}{m^2}\right)^n h(P) + \frac{1}{m^2}\left(1 + \frac{2}{m^2} + ... + \left(\frac{2}{m^2}\right)^n\right)(C_1' + C_2) \tag{6.5}$$

use the formula for infinite geometric series and the fact that $m \geq 2$ we get the following inequality:

$$h(P_n) < \left(\frac{2}{m^2}\right)^n h(P) + \frac{C_1' + C_2}{m^2 - 2} \leq \frac{1}{2^n}h(P) + \frac{1}{2}(C_1' + C_2) \tag{6.6}$$

If we make $n$ sufficiently large, we can make this first term in the righthand side drop to less than 1, so every $P$ is a linear combination of the $Q_r$ and some $P_n$, with $P_n$ less than a finite height independent of $P$. There are only finitely many points with height less than that fixed amount, so the group is generated by those points and the $Q_i$, so it is finitely generated. □

6.2. **Torsion.** The torsion group of an elliptic curve is well-understood over $\mathbb{Q}$, due to the following theorem of Mazur:

**Theorem 6.8** (Mazur,[12])**.** *The torsion group for an elliptic curve over $\mathbb{Q}$ is isomorphic to one of the following fifteen groups:*

$$\mathbb{Z}/m\mathbb{Z} \text{ for } m = 1, 2, ..., 10, 12 \tag{6.7}$$

$$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2m\mathbb{Z} \text{ for } m = 1, 2, 3, 4 \tag{6.8}$$

More recent results have generalized Mazur's original theorem to higher degree extensions of $\mathbb{Q}$. Kamienny's work [10], combined with the earlier work of Kenku and Momose [11], proved the case for quadratic field extensions. Recent work by Derickx and others [4] has given a similar level of

understanding for cubic field extensions. Which torsion groups occur infinitely often is known for extensions of degree up to six [9][5]. However, the question of what torsion groups are possible over higher degree extensions remains open.

For a long time, it was unknown whether or not the number of possible torsion groups could be bounded by a function of the degree. Finally, in 1992 Merel proved the following theorem, which had been known as the Uniform Boundedness Conjecture:

**Theorem 6.9** (Merel,[13])**.** *Let $E$ be an elliptic curve defined over a number field $K$ with degree $d$. Then $E$ cannot possess a point with a prime order higher than $d^{3d^2}$.*

This was improved to an exponential bound by Oesterle, and recent work by Derickx et al. has found tight bounds for $d \leq 6$[3]. There is an expectation that the true bound is polynomial, and perhaps even linear, but the question of the true growth rate of the largest possible prime remains open.

## 7. An Overview of Group Cohomology

It is exceedingly difficult to study the rank of elliptic curves "directly"; that is, much of the progress which has been made towards understanding the rank of elliptic curves arises not from the study of explicit points of infinite order on elliptic curves, but from the study of maps between components of the Mordell-Weil group and simpler adjacent structures.

An indespensible tool in this regard is **group cohomology**, which allows one to associate to some fixed $G$-module $A$ a sequence of meaningful groups whose underlying sets are equivalence classes maps from $G$ into $A$. In this section, we give an axiomatic definition of group cohomology and introduce the concepts needed in the following sections. Our exposition mainly follows [15], [6], and [14].

7.1. **Definition.** A **topological group** is a topological space equipped with a group structure such that multiplication and the taking of inverses are both continuous. Topological groups are homogeneous in the sense that it if one knows a base of open sets around 1, called a **filter base**, one may determine a base open sets around every point $g$ by translating open sets around 1 by $g$. Every group may be viewed as a topological group when equipped with the discrete topology.

A **topological $G$-module** $A$ is a topological abelian group $A$ which admits an additional $G$-set structure, for $G$ a topological group, such that the group action is continuous and for all $g \in G$ and $a, b \in A$, we have $g \cdot (a + b) = g \cdot a + g \cdot b$. Group cohomology was first developed for topological $G$-modules $A$ for which both $G$ and $A$ are equipped with the discrete topology. It turns out, however, that the theory extends well to topological $G$-modules $A$ for which $A$ is discrete and $G$ is **profinite**. To say that $G$ is profinite is to say that it is compact, Hausdorff, and totally disconnected

as a topological space or, equivalently, it is the inverse limit of a system of discrete finite groups. *From now on, we shall assume that $G$ is always profinite and $A$ is always discrete, unless otherwise specified.* We refer to the category of such objects, with morphisms being continuous maps which are $G$-set morphisms and abelian group homomorphisms, as $\mathrm{Mod}_G$.

An **exact sequence** of topological $G$-modules is a sequence of topological $G$-modules $A_i$ and morphisms $\varphi_i : A_i \to A_{i+1}$

$$\cdots \to A_i \xrightarrow{\varphi_i} A_{i+1} \xrightarrow{\varphi_{i+1}} A_{i+2} \to \cdots$$

such that $\operatorname{im}\varphi_i = \ker\varphi_{i+1}$.[1] A **short exact sequence** of topological $G$-modules is an exact sequence of the form

$$0 \to A \to B \to C \to 0.$$

That is, a short exact sequence is a sequence of three topological $G$-modules $A, B, C$ such that $A \hookrightarrow B \twoheadrightarrow C$ such that the image of the injection coincides exactly with the kernel of the surjection.

Suppose we apply to the exact sequence $0 \to A \xrightarrow{\varphi} B \xrightarrow{\psi} C \to 0$ the functor sending a topological $G$-module to its group of $G$-fixpoints. The application of this functor yields an exact sequence

$$0 \to A^G \xrightarrow{\varphi|_{A^G}} B^G \xrightarrow{\psi|_{B^G}} C^G$$

Indeed, it is clear that $\varphi|_{A^G}$ is injective and that $\ker(\psi|_{B^G}) = (\ker(\psi))^G$. Further, we have that $\operatorname{im}(\varphi|_{A^G}) = (\operatorname{im}\varphi)^G$: if $b = \varphi|_{A^G}(a) \in \operatorname{im}(\varphi|_{A^G})$, then $b$ is contained in $\operatorname{im}\varphi$ and for any $g \in G$ we have that $g \cdot b = g \cdot \varphi_{A^G}(a) = \varphi|_{A^G}(g \cdot a) = \varphi|_{A^G}(a) = b$; on the other hand, if $b = \varphi(a) \in (\operatorname{im}\varphi)^G$, then for any $g \in G$ we have $\varphi(a) = g \cdot \varphi(a) = \varphi(g \cdot a)$, and since $\varphi$ is injective, we have that $g \cdot a = a$. Thus, we have that $\operatorname{im}(\varphi|_{A^G}) = (\operatorname{im}(\varphi))^G = (\ker(\psi))^G = \ker(\psi|_{B^G})$, verifying exactness.

However, this exact sequence is not, in general, short exact sequence. That is, while the resulting sequence is still exact, $B^G$ fails in general to surject onto $C^G$.[2] For example, for $p$ a prime, we may take $G$ to be the subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$[3] generated by

$$M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

We take $B = \mathbb{F}_p^2$, equipped with the natural $G$-action, and $A$ to be $\operatorname{span}\{(1,0)\}$, also equipped with the natural $G$-action, which injects into $B$ via the map $A \to B : (n,0) \mapsto (0,n)$. Finally, we take $C = \operatorname{coker}\varphi = \operatorname{span}\{(1,0)\}$, equipped with the natural $G$-action. Since $M$ has the unique eigenvector $(1,0)$, we have that $A = B = C = \operatorname{span}\{(1,0)\}$. Thus, in the induced exact

[1] Note that exact sequences in fact make sense over any category with kernels and cokernels.

[2] Functors of this sort are called **left exact.**

[3] We use $\mathbb{F}_p$ rather than, say $\mathbb{R}$, so that $G$ is finite, and thus profinite. The example works just as well if we use a field like $\mathbb{R}$ except that $G$ would not be profinite, since it would be countable, and we would like to focus on profinite groups for this exposition.

sequence, the injection $0 \to A^G \to B^G$ is an isomorphism and the image of $B^G \to C^G$ must therefore be $0 \subsetneq C^G$, i.e. our sequence fails to be a short exact sequence.

Since it is in many cases desireable for our exact sequences to end in 0, i.e. with the second-to-last nonzero object surjecting onto the last nonzero object, we ask if there is a way to salvage the situation by continuing the exact sequence past $C^G$ in a natural manner—one of the key reasons group cohomology is useful is that it gives us a way to do exactly that.

In particular, we have the following universal property:

**Theorem 7.1.** *There exists an unique (up to isomorphism) sequence of functors $H^i(G, \bullet) : Mod_G \to Ab$ such that the following properties hold:*

(1) *$H^0(G, \bullet)$ is the functor sending a topological $G$-module to its group of $G$-fixpoints.*[4]

(2) *Given an exact sequence $0 \to A \to B \to C \to 0$ of topological $G$-modules, there exists a* **long exact sequence**

$$0 \longrightarrow H^0(G, A) \longrightarrow H^0(G, B) \longrightarrow H^0(G, C) \longrightarrow H^1(G, A) \longrightarrow H^1(G, B) \longrightarrow H^1(G, C) \to \cdots$$

*The maps taking $H^i(G, C) \to H^{i+1}(G, A)$ are called* **connecting morphisms**.

(3) *A morphism of short exact sequences*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & A' & \longrightarrow & B' & \longrightarrow & C' & \longrightarrow & 0 \end{array}$$

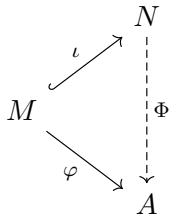
*induces a morphism of long exact sequences*

$$\begin{array}{ccccccccccccc} 0 & \longrightarrow & H^0(G, A) & \longrightarrow & H^0(G, B) & \longrightarrow & H^0(G, C) & \longrightarrow & H^1(G, A) & \longrightarrow & \cdots \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & H^0(G, A') & \longrightarrow & H^0(G, B') & \longrightarrow & H^0(G, C') & \longrightarrow & H^1(G, A') & \longrightarrow & \cdots \end{array}$$

(4) *Suppose $A$ is an* **injective** *topological $G$-module. That is, if $M$ and $N$ are two other topological $G$-modules with an injective morphism $\iota : M \hookrightarrow N$ and morphism $\varphi : M \to A$, then there exists a (not necessarily unique) morphism $\Phi : N \to A$ such that the following*

[4]Letting $H^0(G, \bullet)$ be some left exact functor other than the one sending a topological $G$-module to its group of $G$-fixpoints leads to the theory of **right derived functors**.

*diagram commutes:*




*Then, $H^i(G, A) = 0$ whenever $i > 0$.*

And, indeed, we have the following theorem.

**Theorem 7.2.** *For any topological groups $G$ and $A$, the group $H^i(G, A)$ is always trivial once $i$ is sufficiently large.*

The group $H^n(G, A)$ is often called the $n^{\text{th}}$ cohomology group of $G$ with coefficients in $A$. It turns out that these groups aren't just significant for their sitting in an exact sequence to the right of fixpoint groups. They often have meaningful interpretations in their own right and, in fact, much of the machinery we build in later sections will rely on an important interpretation of the first cohomology groups of elliptic curves qua Galois modules. In this case, the long exact sequence becomes an extremely important tool for reasoning about maps to and from important structures. Thus, we have the following slogan:

> *"Galois cohomology is a tool for extracting long exact sequences from short exact sequences."*

While we will not give a full proof of the above theorem, it is not difficult and often very useful to describe the cohomology groups explicitly; we do so, following Chapter 17 of [6]. Suppose we would like to construct the cohomology groups of a topological $G$-module $A$. We start by considering $C^n(G, A)$, the set of all continuous maps (not necessarily morphisms, just continuous maps) taking $G^n \to A$. The set $C^n(G, A)$ admits a group structure under pointwise addition and elements of $C^n(G, A)$ are often called **$n$-cochains** of $G$ with values in $A$.

Now, we define the **$n^{\text{th}}$ coboundary homomorphisms** $d_n : C^n(G, A) \to C^{n+1}(G, A)$ by

$$\begin{aligned} d_n f(g_1, \ldots, g_{n+1}) = {} & g_1 \cdot f(g_2, \ldots, g_{n+1}) \\ & + \sum_{i=1}^{n} (-1)^i f(g_1, \ldots, g_{i-1}, g_i g_{i+1}, g_{i+2}, \ldots, g_{n+1}) \\ & + (-1)^{n+1} f(g_1, \ldots, g_n). \end{aligned}$$

One can show that $d_n \circ d_{n-1} = 0$ for $n \geq 1$, i.e. the (not necessarily exact) sequence $C^1(G, A) \xrightarrow{d_1} C^2(G, A) \xrightarrow{d_2} \cdots$ is a **cochain complex**. In particular, we have that $\ker d_n \supseteq \operatorname{im} d_{n-1}$ for $n \geq 1$. We define the group

of $n$-**cocycles** $Z^n(G,A) = \ker d_n$ for $n \geq 0$; we also define the group of $n$-**coboundaries** $B^n(G,A)$ as $\operatorname{im} d_{n-1}$ when $n \geq 1$ and in the case $n = 0$, we let $B^0(G,A) = 0$. Now, the $n^{\text{th}}$ cohomology group $H^n(G,A)$ may be defined as $Z^n(G,A)/B^n(G,A)$.

7.2. **Restriction and Galois modules.** Let $G$ a profinite topological group and $H \leq G$ a closed subgroup (we require $H$ to be closed so that it is also profinite). Then, any $G$-module $A$ admits an induced $H$-module structure. Given any $n$-cocycle $f : G^n \to A$ in $C^n(G,A)$, we may consider its restriction $f|_{H^n} : H^n \to A$, and since restrictions of continuous maps are continuous, this restricted map will be an element of $C^n(H,A)$. This induces a **restriction** morphism

$$\operatorname{Res} : H^i(G,A) \to H^i(H,A).$$

Note that this morphism isn't necessarily surjective; indeed, continuous functions on closed subsets do not extend to continuous maps on the ambient space in general.

Let $E$ be an elliptic curve defined over some perfect field $F$. Then, $E$ admits the structure of a $G_F$-module, where here and in the future, we use the notation $G_F$ for $\operatorname{Gal}(\overline{F}/F)$. We consider $G_F$ to be endowed with the **Krull topology**, a filter base of which is given by normal subgroups of finite index (note that these are in bijection with normal extensions of $K$ of finite degree).

We will frequently consider $H^i(G_F, E(\overline{F}))$ and thus we will refer to this group by the shorthand $H^i(F,E)$.

Let $K/\mathbb{Q}$ be a number field. Where $K_v$ is the completion of $K$ at a (not necessarily finite) place $v$, we have a morphism

$$G_{K_v} \hookrightarrow G_K : \sigma \mapsto \sigma|_{\overline{K}}.$$

which is an injection by Krasner's lemma. Then, if $E$ is an elliptic curve defined over $K$, we have the map

$$\operatorname{Res}_v : H^1(K,E) \to H^1(G_{K_v}, E(\overline{K})) \to H^1(K_v, E)$$

where the map $H^1(K,E) \to H^1(G_{K_v}, E(\overline{K}))$ is restriction and $H^1(G_{K_v}, E(\overline{K})) \to H^1(K_v, E)$ is the natural map arising from the inclusion of cochains $C^n(G_{K_v}, E(\overline{K}))$ into $C^n(G_{K_v}, E(\overline{K}_v))$.

## 8. Geometric Machinery

Before returning to the study of elliptic curves properly, we will need a few tools which will help us make sense of the more geometric aspects of the theory of elliptic curves. In particular, we will define curves in general, develop the machinery which allows us to make rigorous sense of "points at infinity", and define divisors on curves. We then develop twists and torsors, along with their relationships to cohomology.

8.1. **Projective Curves and Their Divisors.** An **affine curve** is a set $\{(x,y) \in F^2 : f(x,y) = 0\}$ for some field $F$ that we will often require to be algebraically closed and some polynomial $f$ which is nonconstant and has infinitely many zeroes. The **degree of an affine curve** is the degree of $f$. If $F'$ is a subfield of $F$ and $f \in F'[x,y]$, then the curve in question is said to be an **affine curve defined over** $F'$. For example, the set

$$C = \{(x,y) \in \overline{\mathbb{Q}}^2 : y^2 = x(x-1)(x-2)\}$$

yields a degree 3 affine curve in the plane of algebraic numbers defined over $\mathbb{Q}$. We say that a map $\psi$ between two affine curves $C_1 = \{f_1(x,y) = 0\}$ and $C_2 = \{f_2(x,y) = 0\}$ is a **morphism of affine curves** if it is the restriction of a polynomial $g$ between their respective ambient affine spaces; equivalently, $f_2 = f_1 \circ g$. Suppose that $F'$ is a subfield of $F$; we say that $\psi$ is an $F'$-morphism if $g \in F'[x,y]^2$. We say that two affine curves are $F'$-isomorphic if there exists an $F'$-morphism between them which has a two-sided inverse that is also an $F'$-morphism.

The equation which yields the degree 3 curve $C$ given as an example in the previous paragraph may look familiar—it is a Weierstrass equation that may be associated to an elliptic curve. However, $C$ is not an elliptic curve since it does not have a point at infinity. Indeed, elliptic curves are never affine curves due to the presence of this abstract additional point. Instead, elliptic curves may be considered to be **projective curves** which live in projective space rather than affine space.

For $F$ a field, the **projective space** $F\mathbb{P}^n$ is the quotient of the space of nonzero elements in $F^{n+1}$ under the equivalence relation $[x_1, \ldots, x_{n+1}] \sim [\lambda x_1, \ldots, \lambda x_{n+1}]$ with $0 \neq \lambda \in F$; we may refer simply to $\mathbb{P}^n$ when the field is understood. Note that elements of $F\mathbb{P}^n$ may thus be identified with lines in $F^{n+1}$ which yields the following important interpretation. In $F^{n+1}$, we may consider an affine copy of $F^n$ away from the origin. Then, almost all lines intersect this copy of $\mathbb{F}^n$ at a unique point; these lines may be identified with $\mathbb{F}^n$. All other lines do not intersect the copy at any point and these lines are the "points at infinity". For example, all lines with nonzero slope $\mathbb{R}^2$ intersect the affine copy $\{(x,-1), x \in \mathbb{R}\}$ of $\mathbb{R}^1$ at a unique point, while the line $\{y = 0\}$ does not intersect this copy at all. However, the sequence of points $\{[n,-1]\}_{n\in\mathbb{N}}$ corresponding to the sequence of points $\{(n,-1)\}_{n\in\mathbb{N}}$ in our affine copy, can be seen to approach, in the quotient topology, the point $[1,0]$ which corresponds to the line $\{y = 0\}$; the same is true for the sequence $\{[-n,-1]\}_{n\in\mathbb{N}}$. Thus, we may interpret $\mathbb{RP}^1$ as the real line together with a single additional "point at infinity", which in some sense plays a dual role as positive and negative infinity. More formally, $\mathbb{RP}^1$ is topologically equivalent to the one-point compactification of the real line, namely $S^1$.

A projective curve is a set $\{(x,y,z) \in F\mathbb{P}^2 : f(x,y,z) = 0\}$ for $f$ a homogeneous polynomial. The **degree of a projective curve** is the degree of $f$ and if $f \in F'[x,y,z]$ for $F'$ a subfield of $F$, then the curve in question is said to be a **projective curve defined over** $F'$. We say that a map $\psi$ between

two projective curves $C_1 = \{f_1(x, y, z) = 0\}$ and $C_2 = \{f_2(x, y, z) = 0\}$ is a **morphism of projective curves** if it is the restriction of a polynomial $g$ between their respective projective affine spaces; equivalently, $f_2 = f_1 \circ g$. Suppose that $F'$ is a subfield of $F$; we say that $\psi$ is an $F'$-morphism if $g \in F'[x, y, z]^3$. We say that two projective curves are $F'$-isomorphic if there exists an $F'$-morphism between them which has a two-sided inverse that is also an $F'$-morphism.

The relationship between elliptic curves as defined earlier in this paper and elliptic curves considered as projective curves works as follows. Suppose we have a Weierstrass equation $y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{Q}$ (or more generally any field $L$, in which case we say that $E$ is defined over $L$) such that the cubic on the RHS has distinct roots in $\overline{\mathbb{Q}}$ (or $\overline{L}$). Then, instead of defining an elliptic curve $E$ as $\{y^2 = x^3+Ax+B\}\cup\{O\}$ it is possible to define $E$ as the projective plane curve cut out by the homogeneous polynomial $y^2z = x^3 + Axz^2 + Bz^3$. Note that this has a rational point $[0, 1, 0]$, allowing us to equip this curve a group law, and this point is not on the affine curve cut out by $y^2 = x^3 + Ax + B$. Indeed, we recover our original curve by the substitution $x \mapsto x/z$ and $y \mapsto y/z$ only when $z$ is nonzero, and hence if we would like to work with the affine curve cut out by $y^2 = x^3+Ax+B$, we must replace the point $[0, 1, 0]$ with $O$ the abstract point at infinity. Thankfully, the point $[0, 1, 0]$ is the only data that we lose; it is easy to see that any point of the form $[r, 1, 0]$ on $y^2z = x^3 + Axz^2 + Bz^3$ must have $r = 0$. Thus, the data of the projective curve $\{y^2z = x^3 + Axz^2 + Bz^3\}$ and the affine curve $\{y^2 = x^3 + Ax + B\}$ along with the point $O$ at infinity are essentially the same, at least so far as the group law is concerned.

It turns out that this generalizes to all plane curves. That is, given a plane curve $C = \{f(x, y) = 0\}$ with $f$ of degree $d$, we may "projectivize" $C$ by **homogenizing** $f$, i.e. we lift $C$ to the projective curve $\widehat{C} = \{z^d f(x/z, y/z)\}$. The curve $\widehat{C}$ is called the **projective closure** of $C$. Given $\widehat{C}$, we may return to $C$ by **dehomogenizing**, i.e. taking $z = 1$, during which we may lose information about "points at infinity" where $z = 0$ in $\widehat{C}$.

**Remark 8.1.** *One reason why projective curves are convenient to work with is Bezout's Theorem: let $C_1$ and $C_2$ are projective curves in $K\mathbb{P}^2$, with $K$ an algebraically closed field, given by $f_1$ and $f_2$, respectively; if $C_1$ and $C_2$ do not intersect at infinitely many points (equivalently, the greatest common divisor of $f_1$ and $f_2$ is constant), then the number of intersection points of $C_1$ and $C_2$ in $K\mathbb{P}^2$ is exactly $\deg(f_1)\deg(f_2)$ when counted with multiplicity. This fails in the affine setting, where the number of intersection points is at most $\deg(f_1)\deg(f_2)$ when counted with multiplicity. For example, the curves cut out by $f_1(x, y) = y - x$ and $f_2(x, y) = y - x - 1$ in $\mathbb{C}^2$ have no intersection points, but the corresponding curves in $\mathbb{CP}^2$ cut out by $zf_1(x/z, y/z) = y - x$ and $zf_2(x/z, y/z)$ in $\mathbb{CP}^2$ can easily be seen to intersect at exactly one point, namely $[1, 1, 0]$.*

From now on, unless otherwise specified, by "curve" we will refer to projective curves in $\overline{\mathbb{Q}}$ defined over $\mathbb{Q}$. If we specify a curve by an equation in two variables, we refer to the projective closure of the affine curve cut out by the equation in question; if we refer to points $(x, y)$, we implicitly mean $[x, y, 1]$.

To a curve $C$, we may associate **divisors**, which are formal linear combinations of finite collections of points on $C$. That is, a divisor is $\sum_{P \in C} n_P P$ where all but finitely many $n_P$ are zero. For example, if $C$ is given by $4x^2 - x^3 - 2 = 0$, then $D = 2(-\sqrt[3]{2}, 0) + 3(0, \sqrt{1/2})$ is a divisor on $C$.

**Remark 8.2.** *Divisors are formal linear combinations; one should not confuse + in this case with addition on a group law which may be defined on the curve in question.*

Given a divisor $D$, we may define its **degree** $\deg D$ as follows. For a point $P$ of $C$, define $\varphi(P) = \min_{K \ni P, [K:\mathbb{Q}] < \infty} [K : \mathbb{Q}]$. Then, define $\deg(\sum_{P \in C} n_P P) = \sum_{P \in C} n_P \varphi(P)$. In other words, we define the degree of a single point $P$ to be the degree of the minimal number field $K$ such that $K^2 \ni P$ and we extend linearly.

For example, the degree of the divisor $D = 2(-\sqrt[3]{2}, 0) - 3(0, \sqrt{1/2})$ on the curve cut out by $y^2 - x^3 - 2 = 0$ is 0 since the splitting field of $x^3 + 2$ is of degree 3 over $\mathbb{Q}$ and the splitting field of $x^2 - \frac{1}{2}$ is of degree 2; thus, extending linearly yields:

$$\overbrace{\overbrace{2\overbrace{(-\sqrt[3]{2}, 0)}^{3}}^{6} - 3\overbrace{(\sqrt{1/2}, 0)}^{2}}^{-6}{}^{0}.$$

Let $C$ be a curve. Consider the field of rational functions $f(x, y)$ in two variables on $C$ which are ratios of homogeneous polynomials of the same degree. These may be written *locally* as $f(t)$ a rational function in one variable. For any point $P \in C$, we have that if $f(x, y)$ may be represented as $f(t)$ around $P$, then $f(t)$ may be written as $(t - P)^\nu \widehat{f}$ for some nonzero rational function $\widehat{f}$ such that neither the numerator nor denominator of $\widehat{f}$ contains a factor $(t - P)$. The integer $\nu$ is called the **order** $\mathrm{ord}_P(f)$ **of** $f$ **at** $P$. For a given $f$, there are only finitely many points $P \in C$ at which $\mathrm{ord}_P(f) \neq 0$, and thus $\mathrm{div} : f \mapsto \sum_{P \in C} \mathrm{ord}_P(f) P$ is a map taking rational functions on $C$ to divisors on $C$. A divisor $D$ is said to be **principal** if there is a rational function $f$ on $C$ with $\mathrm{div}(f) = D$. It turns out that the degree of any principal divisor is zero. Two divisors are said to be **linearly equivalent** if their difference is principal; note that this partitions divisors into equivalence classes over which degree is well-defined.

8.2. **Twists and Torsors.** In our overview of group cohomology, we promised that cohomology groups would often have meaningful interpretations; the

time has come for us to make due on that promise. Our exposition in this subsection follows [15].

We begin with a rather imprecise premise which we will not attempt to make completely rigorous. Suppose we have a category $C$ of objects which are "defined over" a perfect field $K$, in some well-specified sense. Further, suppose that for every extension $K'/K$, there exists a category of objects $C_{K'}$ which are defined "defined over" $K'$ and that for every object $V$ of $C$, there exists a corresponding object of $V_{K'}$ in $C_{K'}$, i.e. given an object $V$ defined over $K$, we can "base change" to obtain an object $V_{K'}$ over $K'$. We say two objects $V$ and $W$ in $C$ are **twists** of one another if there exists an isomorphism between $V_{\overline{K}}$ and $W_{\overline{K}}$ in $C_{\overline{K}}$. We will often identify twists in $C$ if they are isomorphic to one another in $C$.

For example, the category of elliptic curves defined over $\mathbb{Q}$ along with $\mathbb{Q}$-morphisms satisfies the properties we require of $C$ (although the machinery we are currently developing will not be applied directly to this category in this and the following sections). It can be shown that the elliptic curves given by $y^2 = x^3+1$ and $2y^2 = x^3+1$ are not $\mathbb{Q}$-isomorphic, but we may transform the former into the latter via the change of variables $(x, y) \mapsto (x, y\sqrt{2})$ in $\mathbb{Q}(\sqrt{2})$; thus, the two elliptic curves are twists of one another. Note that elliptic curves are twists of one another if and only if they have the same $j$-invariant.

For some object $V$ in our category $C$ of objects over $K$ satisfying the vaguely set forth principles above, let $\mathrm{Aut}(V_{\overline{K}})$ be the set of automorphisms of $V_{\overline{K}}$ equipped with an action of $G_K$. If $\mathrm{Aut}(V_{\overline{K}})$ is abelian, then it admits the structure of a topological $G_K$-module and thus we may speak of its cohomology groups; if not, we can still make sense of its cohomology via pointed sets rather than groups (we will not elaborate on this point since $\mathrm{Aut}(V_{\overline{K}})$ will be abelian in all the situations we consider in this exposition). In either case, we have an injection

$$\{\text{twists of } V\}/\cong_K \hookrightarrow H^1(G_K, \mathrm{Aut}(V_{\overline{K}})),$$

which will in many situations be a bijection, given as follows. If $W$ is a twist of $V$ in $C$, then we choose an isomorphism $\varphi : W_{\overline{K}} \to V_{\overline{K}}$. Then the map $G \to \mathrm{Aut}(V_{\overline{K}}) : g \mapsto (g \cdot \varphi) \circ \varphi^{-1}$ is a 1-cocycle and yields an element of $H^1(G_K, \mathrm{Aut}(V_{\overline{K}}))$.

Thus, we have the following imprecise slogan:

> *"The group or pointed set $H^1(G_K, \mathrm{Aut}(V_{\overline{K}}))$ classifies twists of $V$."*

This principle will turn out to be essential to our understanding of elliptic curves via its application to torsors, which we now define.

Let $E$ be an elliptic curve defined over a perfect field $K$; then, we have a group action $E \times E \to E$ which is (1) free and (2) transitive. Furthermore, this map is (3) always the restriction of a homogeneous polynomial map

in $\mathbb{Q}[x, y, z, w, v]^3$ taking $K\mathbb{P}^4 \to K\mathbb{P}^2$.[5] More generally, we may consider curves $C$ which admit a group action $E \times E \to E$ satisfying these three properties. Such curves $C$ along with such group actions $E \times C \to C$ are known as **$K$-torsors** of $E$. A morphism of $K$-torsors $C_1$ and $C_2$ of $E$ is a $K$-morphism $\varphi$ of curves which is also $E$-invariant, i.e. for any $P \in E$ and $Q \in C_1$, we have $\varphi(P + Q) = P + \varphi(Q)$.

We note that if $K'/K$ is a field extension of $K$, any $K$-torsor of $E$ may be considered as a $K'$-torsor, and thus we can make sense of twists of $K'$-torsors of $E$. We would like to use cohomology to classify twists of $E$ considered as a $\mathbb{Q}$-torsor of $E$.

In this specific instance, we are especially fortunate. Our method of using $H^1$ to classify twists of $E$ as a $K$-torsor of $E$ is particularly powerful for the following reasons:

(1) It is nontrivial, but nonetheless true, that the $K$-torsors of $E$ are exactly the twists of $E$ as a $K$-torsor, i.e. all $K$-torsors of $E$ become isomorphic to $E$ as a $K$-torsor of $E$ over $\overline{K}$. Thus, using $H^1$ to classify twists of $E$ as a $K$-torsor of $E$ yields a classification of all $K$-torsors of $E$.
(2) The automorphisms of $E$ as a $\overline{K}$-torsor of $E$ are the maps $+Q : P \mapsto P + Q$. In particular, the automorphism group is abelian and isomorphic to $E$ as a group, so $H^1(G_K, \mathrm{Aut}(E_{\overline{K}})) \cong H^1(K, E)$.
(3) We noted earlier that the injection of twists into $H^1$ would often be a bijection. In this case, it is indeed a bijection.

Thus, we have

$$\{K\text{-torsors of } E\}/\cong_K \longleftrightarrow H^1(K, E).$$

Furthermore, we have the following theorem.

**Theorem 8.3.** *Let $C$ be a $K$-torsor of $E$. The following are equivalent:*

(1) *$C$ is a* **trivial $K$-torsor of** *$E$, i.e. $C \cong E$ as $K$-torsors of $E$.*
(2) *$C(K) \neq \{\}$, i.e. $C$ has a $K$-rational point.*
(3) *$C$ corresponds to $0$ under the above bijection.*

That is, nontrivial elements of $H^1$ correspond to $K$-torsors of $E$ (up to $K$-isomorphism) which do not have $K$-points. Thus, using this correspondence, we may apply the Galois cohomological tools developed in the previous section to understand $K$-torsors of $E$. We do so in the following section.

## 9. The Shafarevich-Tate and Selmer Groups

Using the machinery developed in the previous sections, we define important groups, $Ш(E)$ and $\mathrm{Sel}^n(E)$, associated to an elliptic curve $E$, which, in addition to being interesting for a number of reasons in their own rights,

[5]For readers who have experience with varieties, this is to say that this map is a $K$-morphism from the product variety $E \times E$ to the elliptic curve $E$.

sit in an exact sequence with a quotient group of $E(\mathbb{Q})$, allowing us to obtain a bound on rank indirectly by studying these adjacent structures. Our exposition in this section partially follows [15].

We have for an elliptic curve $E$ defined over $\mathbb{Q}$ that $H^1(\mathbb{Q}, E)$ classifies $\mathbb{Q}$-torsors of $E$ and that the nontrivial elements of this group correspond to $\mathbb{Q}$-torsors of $E$ with no $\mathbb{Q}$-points. Under this interpretation, the map $\mathrm{Res}_v$ lifts a torsor into $\mathbb{Q}_v$ in the natural way. We define the **Tate-Shafarevich group** Ш by

$$Ш(E) = \bigcap_{v \leq \infty} \ker \left[ H^1(\mathbb{Q}, E) \xrightarrow{\mathrm{Res}_v} H^1(\mathbb{Q}_v, E) \right]$$

Note that $v$ ranges over all places of $\mathbb{Q}$, finite or infinite. That is, elements of the Tate-Shafarevich group are nontrivial $\mathbb{Q}$-torsors of $E$ which become trivial $\mathbb{Q}_v$ for all places $v$.

Note the connection with the Hasse-Minkowski theorem here: if $Ш(E)$ were always trivial, the Hasse-Minkowski theorem would hold for elliptic curves. However, the group $Ш(E)$ is often not trivial; indeed, it is not even known that $Ш(E)$ is always finite, although this is conjectured to be true, and there is no known method to compute $Ш(E)$ in general.

The group of $n$-torsion points of $Ш(E)$, denoted $Ш(E)[n]$ sit in an exact sequence with the group $\mathrm{Sel}^n(E)$, which we now define. Note that the formal definition is somewhat abstract, but $\mathrm{Sel}^n(E)$ does have a geometric interpretation which we will state later.

Let $E[n]$ denote the $n$-torsion points $E$. Where $[n]$ is the multiplication-by-$n$ map, we have the following exact sequence:

$$0 \longrightarrow E[n] \longrightarrow E \xrightarrow{[n]} E \longrightarrow 0.$$

Using cohomology, we extract from this short exact sequence the long exact sequence

$$\begin{aligned} 0 \longrightarrow H^0(G_{\mathbb{Q}}, E[n]) \longrightarrow H^0(\mathbb{Q}, E) \xrightarrow{[n]} H^0(\mathbb{Q}, E) \\ \hookrightarrow H^1(G_{\mathbb{Q}}, E[n]) \longrightarrow H^1(\mathbb{Q}, E) \xrightarrow{[n]} H^1(\mathbb{Q}, E) \to \cdots \end{aligned}$$

or, remembering that $H^0(G, \bullet)$ is the fixpoint functor,

$$\begin{aligned} 0 \longrightarrow E(\mathbb{Q})[n] \longrightarrow E(\mathbb{Q}) \xrightarrow{[n]} E(\mathbb{Q}) \\ \hookrightarrow H^1(G_{\mathbb{Q}}, E[n]) \longrightarrow H^1(\mathbb{Q}, E) \xrightarrow{[n]} H^1(\mathbb{Q}, E) \to \cdots . \end{aligned}$$

We may shorten the subsequence of the first seven terms (including 0) to the top row of the following diagram; we may repeat this process with $\mathbb{Q}$

replaced by $\mathbb{Q}_v$ for any place $v$, which yields the bottom row:

(9.1)

$$\begin{array}{ccccccccc} 0 & \longrightarrow & E(\mathbb{Q})/nE(\mathbb{Q}) & \longrightarrow & H^1(G_{\mathbb{Q}}, E[n]) & \longrightarrow & H^1(\mathbb{Q}, E)[n] & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \prod_{v\leq\infty} \text{incl}} & & \downarrow{\scriptstyle \prod_{v\leq\infty} \text{Res}_v} & \searrow^{\tau} & \downarrow{\scriptstyle \prod_{v\leq\infty} \text{Res}_v} & & \\ 0 & \longrightarrow & \prod_{v\leq\infty} E(\mathbb{Q}_v)/nE(\mathbb{Q}_v) & \longrightarrow & \prod_{v\leq\infty} H^1(G_{\mathbb{Q}_v}, E[n]) & \longrightarrow & \prod_{v\leq\infty} H^1(\mathbb{Q}_v, E)[n] & \longrightarrow & 0 \end{array}$$

We now define the $n$-**Selmer group** $\text{Sel}^n(E)$ by $\ker(\tau)$. From 9.1, we have the following diagram:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & E(\mathbb{Q})/nE(\mathbb{Q}) & \longrightarrow & H^1(G_{\mathbb{Q}}, E[n]) & \longrightarrow & H^1(\mathbb{Q}, E)[n] & \longrightarrow & 0 \\ & & \downarrow & & \downarrow{\scriptstyle \tau} & & \downarrow{\scriptstyle \prod_{v\leq\infty} \text{Res}_v} & & \\ 0 & \longrightarrow & 0 & \longrightarrow & \prod_{v\leq\infty} H^1(\mathbb{Q}_v, E)[n] & = & \prod_{v\leq\infty} H^1(\mathbb{Q}_v, E)[n] & \longrightarrow & 0. \end{array}$$

From here, the Ker-Coker Sequence yields the exact sequence

$$0 \to E(\mathbb{Q})/nE(\mathbb{Q}) \to \text{Sel}^n(E) \to Ш(E)[n] \to 0, \tag{9.2}$$

i.e. $\text{Sel}^n(E)$ surjects onto $Ш(E)[n]$ with kernel $E(\mathbb{Q})/nE(\mathbb{Q})$.

The group $\text{Sel}^n(E)$ has the following geometric interpretation: nontrivial elements of $\text{Sel}^n(E)$ are pairs $([C], [D])$ where $[C]$ is an isomorphism class of nontrivial $\mathbb{Q}$-torsors of $E$, trivial in $\mathbb{Q}_v$ for every $v$, and $[D]$ is the linear equivalence class of the degree $n$ divisor $D$ on $C$. It is nonobvious, but nonetheless true, that any element $C$ of $Ш(E)[n]$ must gain a point $P$ in some number field $L/\mathbb{Q}$ with $[L : \mathbb{Q}] = n$ and thus we may lift $C$ to the element $([C], [P])$ in $\text{Sel}^n(E)$. Thus, $\text{Sel}^n(E)$ surjects onto $Ш(E)[n]$, but the kernel of this surjection is not (necessarily) trivial, but instead equal to $E(\mathbb{Q})/nE(\mathbb{Q})$ as in 9.2.

The exact sequence 9.2 is extremely useful since it turns out that $\text{Sel}^n(E)$, unlike $Ш(E)$, is always finite and, in general, much more is known about $\text{Sel}^n(E)$ than $Ш(E)$. Thus, we may obtain information on the size of $E(\mathbb{Q})/nE(\mathbb{Q})$ andthus the rank of $E(\mathbb{Q})$ by studying $\text{Sel}^n(E)$, one method of which we will elaborate upon in the next section.

## 10. Quartic Forms and the Average Rank of Elliptic Curves

In their celebrated 2015 paper [1], Manjul Bhargava and Arul Shankar give a bound on the limit superior of the average rank of elliptic curves via a count of binary quadratic forms. In this section, we give a very brief summary of their strategy and state their results.

Let $E$ be an elliptic curve defined over $\mathbb{Q}$. Mordell's Theorem implies that $E(\mathbb{Q}) \cong_{\text{Ab}} \mathbb{Z}^r \times T$ for $T$ a torsion group; the integer $r$ is called the **rank** of $E$. We would like to understand the average rank of elliptic curves. Since the number of elliptic curves is infinite, we must specify how elliptic curves are ordered in order to understand the average value of rank. All

elliptic curves over $\mathbb{Q}$ are $\mathbb{Q}$-isomorphic to a unique curve $E_{A,B}$ cut out by $y^2 = x^3 + Ax + B$ such that $A$ and $B$ are integers and $p^4 \mid A \Rightarrow p^6 \nmid B$. We define the **height** $h$ of $E_{A,B}$ as $h(E_{A,B}) = \max\{|A|^3, B^2\}$. This yields an ordering on elliptic curves allowing us to make sense of average rank. It is still not known whether the average rank of elliptic curves exists, but Bhargava and Shankar obtain a bound on the limit superior of the average rank of elliptic curves. In particular, they show:

**Theorem 10.1.** *The limit superior of the average rank of elliptic curves ordered by height is not greater than* 1.5.

Since any elliptic curve $E$ over $\mathbb{Q}$ is isomorphic as an abelian group to $\mathbb{Z}^r \times T$, we have that $E(\mathbb{Q})/2E(\mathbb{Q})$ is isomorphic to $(\mathbb{Z}/2)^r \times (T/2)$. Thus, using 9.2, we have that $2^r \leq |\operatorname{Sel}^2(E)|$, i.e. if we can calculate the average size of $\operatorname{Sel}^2(E)$, we have a bound on the limit superior of the average rank of elliptic curves over $\mathbb{Q}$.

How do we reason about the average size of $\operatorname{Sel}^2(E)$? It turns out that elements of $\operatorname{Sel}^2(E)$ admit another geometric interpretation in terms of locally soluble 2-coverings. A **locally soluble 2-covering** of $E$ is a curve $C$ whose $\mathbb{C}$-points are homeomorphic to a 2-torus, along with a $\mathbb{C}$-isomorphism $\varphi : C \to E$ and a $\mathbb{Q}$-morphism $\psi : C \to E$ such that

$$\begin{array}{ccc} C & & \\ {\scriptstyle\varphi}\downarrow & \searrow^{\psi} & \\ E & \xrightarrow[P\mapsto P+P]{} & E \end{array}$$

commutes and $C$ has a point in $\mathbb{Q}_v$ for every place $v$. Two locally soluble 2-coverings $C_1$ and $C_2$ with respective isomorphisms $\varphi_1$ and $\varphi_2$ are said to be isomorphic if there exists a $\mathbb{Q}$-isomorphism of curves $\Phi : C_1 \to C_2$ along with a point $Q \in E[2]$ such that the following diagram commutes:

$$\begin{array}{ccc} E & \xrightarrow{P\mapsto P+Q} & E \\ {\scriptstyle\varphi_1}\uparrow & & \uparrow{\scriptstyle\varphi_2} \\ C_1 & \xrightarrow[\Phi]{} & C_2 \end{array}$$

Then, we may view elements of $\operatorname{Sel}^2(E)$ as being isomorphism classes of locally soluble 2-coverings of $E$. If a locally soluble 2-covering has a rational point, it is said to be a **soluble 2-covering**. Isomorphism classes of soluble 2-coverings correspond exactly to elements of $E(\mathbb{Q})/2E(\mathbb{Q})$ inside $\operatorname{Sel}^2(E)$.

Now, a result by Birch and Swinnerton-Dyer [18] states that any locally soluble 2-covering is isomorphic to one cut out by

$$y^2 = ax^4 + bx^3 + cx^2 + dx + e$$

for $a, b, c, d, e \in \mathbb{Q}$. Thus, we obtain a binary quartic form

$$ax^4 + bx^3y + cx^2y^2 + dxy^3 + ey^4$$

by taking the homogenization of the RHS.

Recall that binary quartic forms have two invariants $I$ and $J$. It turns out that the binary quartic forms $f$ obtained by the above process when our elliptic curve $E$ is given by $y^2 = x^3 + Ax + B$ have $I(f) = A$ and $J(f) = B$. More specifically, we have an injection

$$\mathrm{Sel}^2(E) \hookrightarrow \{\text{binary quartic forms } f | I(f) = A, J(f) = B\}/\mathrm{SL}_2(\mathbb{Q}) \cdot \mathbb{Q}^\times$$

(For various technical reasons, Bhargava and Shankar count forms up to projective linear equivalence, hence the appearence of $\mathbb{Q}^\times$). Furthermore, if we require that $f$ satisfy additional minor technical conditions, the injection becomes a bijection.

Bhargava and Shankar employ a highly technical geometry of numbers argument, in the spirit of those of Minkowski but significantly more involved, to deduce that the average size of the appropriate set of forms is 3, when we order by height. Thus, we have that the average size of $\mathrm{Sel}^2(E)$ is 3, which immediately yields an upper bound of 1.5 on the limit superior of the average rank of elliptic curves.

## Acknowledgements

We would first like to thank Dan Dore and Libby Taylor for mentoring this project. We would also like to thank Brian Conrad and Ravi Vakil for their insights and helpful conversations. Finally, we would like to thank Chris Ohrt and the rest of the SURIM participants for making this summer mathematically inspiring and exciting.

## References

[1] M. Bhargava and A. Shankar. Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves. *Annals of Mathematics*, 181(1):191–242, 2015.

[2] H. Davenport. *Multiplicative Number Theory*. Springer-Verlag, 1980.

[3] M. Derickx, S. Kamienny, W. Stein, and M. Stoll. Torsion points on elliptic curves over number fields of small degree. *Preprint*, 07 2017.

[4] M. Derickx and F. Namjan. Torsion of elliptic curves over cyclic cubic fields. *Mathematics of Computation*, 88:2443–2459, 2019.

[5] M. Derickx and A.V. Sutherland. Torsion subgroups of elliptic curves over quintic and sextic number fields. *Proceedings of the American Mathematical Society*, 145:4233–4245, 2017.

[6] D.S. Dummit and R.M. Foote. *Abstract Algebra*. Wiley, 2003.

[7] C. F. Gauss and A. A. Clarke. *Disquisitiones Arithmeticae*. Yale University Press, 1965.

[8] D. Hilbert. *Theory of Algebraic Invariants*. Cambridge University Press, 1993.

[9] D. Jeon, C. H. Kim, and E. Park. On the torsion of elliptic curves over quartic number fields. *Journal of the London Mathematical Society*, 74:1–12, 2006.

[10] S. Kamienny. Torsion points on elliptic curves and q-coefficients of modular forms. *Inventiones Mathimaticae*, 109:221–229, 1992.
[11] M.A. Kenku and F. Momose. Torsion points on elliptic curves defined over quadratic fields. *Nagoya Mathematical Journal*, 109:125–149, 1988.
[12] B. Mazur. Rational isogenies of prime degree. *Inventiones Mathimaticae*, 44:129–162, 1978.
[13] L. Merel. Bornes pour la torsion des courbes elliptiques sur les corps de nombres. *Inventiones Mathimaticae*, 124:437–449, 1996.
[14] Park City Mathematics Institute. *Galois Cohomology*, 1991.
[15] B. Poonen. The selmer group, the shaferavich-tate group, and the weak mordell-weil group. 2002.
[16] J.P. Serre. *A Course in Arithmetic*. Springer-Verlag, 1973.
[17] J.H. Silverman. *The Arithmetic of Elliptic Curves*. Springer, 1986.
[18] Birch B.J. Swinnerton-Dyer, H.P.F. Notes on elliptic curves. i. *Journal für die reine und angewandte Mathematik*, 212:7–25, 1963.

*E-mail address*: yuzu@stanford.edu

*E-mail address*: chainpur@stanford.edu

*E-mail address*: mcs0042@stanford.edu