History of the Birch and Swinnerton-Dyer Conjecture

James Bourke

Thesis submitted for the degree of M.Sc

Department of Mathematics and Statistics
Maynooth University

September 2023

Head of Department: Prof. Stephen Buckley

Supervisor: Dr. Detta Dickinson

Abstract

In this thesis, an account of the Birch and Swinnerton-Dyer conjecture and some of its history are presented. A framework on the theory of elliptic curves is developed from the ground up. Then, several advanced definitions and results from algebraic geometry are given. This includes an accessible chapter about the Tate-Shafarevich group. These results will help illuminate some of the reasoning behind Birch and Swinnerton-Dyer's work. Finally, a thorough, up to date report on the progress of the conjecture is included.

Acknowledgements

First and foremost, I would like to thank my mother Miriam, my father John, my brother Eoin and my dogs Buddy and Ruby. You have not only provided everything for me, but have also been a constant source of love, encouragement, joy and laughter. You have supported me through thick and thin, for which I am eternally grateful. I would also like to thank my grandparents Kathleen and James for their continuous love and support.

I would also like to thank all my friends. You have all been tremendous for as long as I have known you. You never fail to make me laugh, and never fail to help me when I ask for it.

Finally, I would like to thank everyone at the Department of Mathematics and Statistics, but none so as much as Dr. Detta Dickinson. Without her, this thesis would not have been possible. She has been an outstanding mentor and teacher, and is without doubt one of the greatest mathematicians I have ever met. Her advice and support have been invaluable, and for this I am forever indebted to her.

Contents

Notation

$:=$	Defined to be.
$\cong$	Isomorphic to.
S^1	Circle group.
$\mathrm{R}[x_1, \ldots, x_n]$	Ring of polynomials in n variables defined over the ring R.
$\mathrm{R}(x_1, \ldots, x_n)$	Field of rational functions in n variables defined over the ring R.
Δ	Discriminant of an elliptic curve.
$+$	Group operation for abelian groups.
$\mathrm{Char}(\mathbb{F})$	Characteristic of the field $\mathbb{F}$.
$\mathbb{Z}^r$	r-fold Cartesian product of $\mathbb{Z}$.
gcd	Greatest common divisor.
$\mathbb{Q}^*$	Multiplicative group of non-zero rational numbers.
$\mathrm{Ker}(\phi)$	Kernel of the homomorphism ϕ.
$\mathrm{Im}(\phi)$	Image of the homomorphism ϕ.
ν_p	p-adic valuation.
$\lvert \cdot \rvert$	Cardinality / Absolute value.
$\overline{\mathbb{F}}$	Algebraic closure of the field $\mathbb{F}$.
$\mathrm{Gal}(\mathbb{L}/\mathbb{F})$	Galois group of the field extension $\mathbb{L}/\mathbb{F}$.
$A/\sim$	Equivalence classes of the set A under the equivalence relation $\sim$.
$f\vert_A$	Restriction of the function f to the set A.
$Ш_E$	Tate-Shafarevich group of E.
S_n	n-th Selmer group.
$\mathbb{P}_n(\mathbb{F})$	Projective n-space over the field $\mathbb{F}$.
$\circ$	Composition of functions.
det	Determinant.
$n \vert m$	n divides m.
$n \nmid m$	n does not divide m.
$\sqrt[3]{x}$	Cubed root of x.
$\exp(x)$	Exponential of x.
$\Re(s), \Im(s)$	Real and imaginary components of s.
$\Gamma(s)$	Gamma function.
$\mathrm{PSL}_2(\mathbb{Z})$	Modular group / Projective linear group.

Chapter 1

Introduction

For nearly 60 years, the Birch and Swinnerton-Dyer conjecture has captured the attention of mathematicians as one of the deepest, most difficult problems the subject has to offer. The conjecture itself is a culmination of countless results from algebra, number theory and analysis. The interplay of these three fields is evident in the conjecture and plays a significant part in why so many mathematicians are fascinated by it.

In the second chapter, an overview of elliptic curves and an account of why they were developed is included. This includes a comprehensive section on Mordell's theorem. The purpose of this chapter is two-fold. Firstly, to provide motivation to readers and secondly, to establish a good starting point from which the theory underlying the Birch and Swinnerton-Dyer conjecture may be developed.

The purpose of the third chapter is to introduce the Tate-Shafarevich group. This group plays a huge role, not only in Birch and Swinnerton-Dyer's work, but in modern algebraic geometry. It is an intriguing object which has significant implications, particularly in relation to Hasse's famous local-global principle.

After this, a summary of Birch and Swinnerton-Dyer's paper *Notes on elliptic curves I* is given in chapter four. This chapter is included so that the reader may gain some insight into how Birch and Swinnerton-Dyer were able to arrive at their conjecture. This chapter is essentially self-contained and is included for the sake of completeness.

The goal of the fifth chapter is to provide some preliminary notions, not included in previous chapters, that are required to understand the statement of the conjecture and its strong form. It also contains several concepts that are needed to understand the progression of the conjecture and its partial resolutions.

The final chapter includes a complete statement of the conjecture, a list of consequences if the conjecture should turn out to be correct and a history of special cases in which it has been resolved.

There are two appendices, one on projective geometry, which is needed to formalise 'points at infinity' and another on p-adic numbers which includes a statement of Hensel's lemma.

Chapter 2

Elliptic Curves (A Complex History)

From acoustics to optics, and from the quantum world to the solar system, the science of our universe is heavily intertwined with ellipses. In fact, one of the classical systems in physics, a 2D harmonic potential, only admits elliptical paths[1]. For these and other reasons, many of history's greatest mathematicians including Legendre, Jacobi and Weierstrass, sought to understand them.

Though it would not be difficult to argue that ellipses are no more than compressed circles, shapes that have been well understood since antiquity, the study of circles is trivial due to their symmetry. However, we shall discover that because of their eccentric behaviour, studying ellipses gives rise to applications well beyond geometry.

Ellipses are circles that have been scaled (stretched or contracted) horizontally and/or vertically. To describe them algebraically, the x and y variables in the equation for the unit circle can be suitably scaled. An ellipse of width $2a$ and height $2b$ is given by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

It is immediate from this description that the parametrisation of such an ellipse is

$$p(\theta) = (a\cos(\theta), b\sin(\theta)), \text{ where } \theta \in [0, 2\pi).$$

[1] Under the assumption that the paths are stable and closed.

With this parametrisation, it is natural to try and find a formula for the perimeter of an ellipse. Apart from being a nice mathematical curiosity, finding such a formula has far reaching consequences throughout physics. Generally, the stable, periodic motion of one celestial body relative to another follows an elliptical path. For example, the motion of the Earth about the Sun. Finding the perimeter of such an ellipse gives the length of the path the orbiting object needs to travel before it returns to its initial position.

By applying the parametrisation above and the arc length integral,

$$L = \int_a^b |f'(t)|dt$$

the perimeter of an ellipse can be written as

$$P = \int_0^{2\pi} \sqrt{a^2 \sin(\theta)^2 + b^2 \cos(\theta)^2} d\theta.$$

Unfortunately, this integral is notoriously difficult. Attempting to compute it was what led mathematicians to investigate a large family of similar integrals, known as *elliptic integrals* for their historic ties to ellipses.

First note that by rewriting the perimeter integral in Cartesian co-ordinates, it is easier to exploit the vertical and horizontal symmetries of an ellipse. The substitution $x = \sin(\theta)$ implies

$$P = 4 \int_0^1 \frac{\sqrt{a^2x^2 + b^2(1 - x^2)}}{\sqrt{1 - x^2}} dx.$$

Usually, the variable $k := \sqrt{1 - \left(\frac{a}{b}\right)^2}$ is introduced to make this integral simpler. The variable k is known as the *eccentricity* of the ellipse and gives an indication of how elongated it is. The formula then reduces to

$$P = 4b \int_0^1 \frac{\sqrt{1 - k^2x^2}}{\sqrt{1 - x^2}} dx.$$

The standard notation for the above integral is

$$E(k) := \int_0^1 \frac{\sqrt{1 - k^2x^2}}{\sqrt{1 - x^2}} dx,$$

and it is often generalised by allowing the upper limit of integration to vary,

$$E(t,k) := \int_0^t \frac{\sqrt{1-k^2x^2}}{\sqrt{1-x^2}}dx.$$

By allowing k to vary, these integrals form a family known as the *elliptic integrals of the second kind*[2]. They are essential for determining the arc length between arbitrary points on an ellipse. As for *elliptic integrals of the first kind*, they are of the form

$$K(t,k) = \int_0^t \frac{1}{\sqrt{1-x^2}\sqrt{1-k^2x^2}}dx.$$

Whereas $E(t,k)$ gave an indication of how far an object travels along a given path, $K(t,k)$ gives an indication of how long this journey will take. For example, it is used to calculate the period of a simple pendulum. It can also be interpreted as a generalisation of arcsin since if $k = 0$, $K(t,0) = \arcsin(t)$. The same reasoning applies to the inverse of $K(t,k)$ with respect to t (k fixed). This inverse is denoted $t(K,k)$ and can be viewed as a generalisation of sin.

Legendre and Jacobi studied these functions meticulously, as they appear in many classical mechanics problems. While researching the properties of these functions, Jacobi discovered that if the domain of the variable K is extended from $\mathbb{R}$ to $\mathbb{C}$, the functions $t(K,k)$ become *doubly periodic* over $\mathbb{C}$.

Definition 2.1. *A function $f : \mathbb{C} \longrightarrow \mathbb{C}$ is **doubly periodic** if there exist $\omega_1, \omega_2 \in \mathbb{C}$, linearly independent over $\mathbb{R}$, such that*

$$f(z) = f(z+\omega_1) = f(z+\omega_2) \text{ for all } z \in \mathbb{C}.$$

Equivalently, if $L(\omega_1,\omega_2)$ is defined to be the lattice in $\mathbb{C}$ generated by ω_1 and ω_2, i.e.

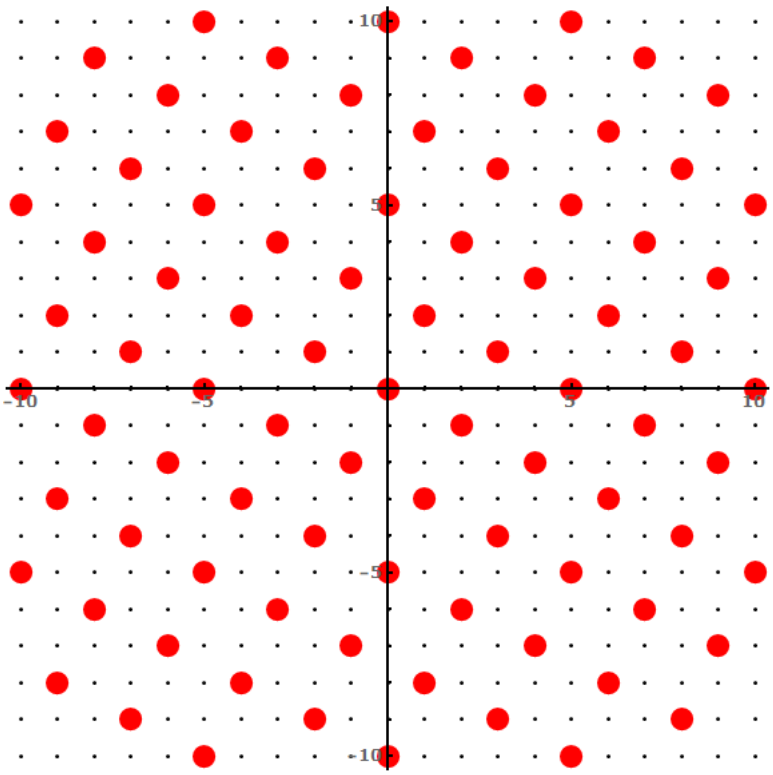
$$L(\omega_1,\omega_2) := \{n\omega_1 + m\omega_2 \mid n,m \in \mathbb{Z}\},$$

then another way of saying f is doubly periodic, with respect to $L(\omega_1,\omega_2)$, is

$$f(z) = f(z+\lambda) \text{ for all } z \in \mathbb{C} \text{ and all } \lambda \in L(\omega_1,\omega_2).$$

[2]The reason they are known as the *second kind* is due to Legendre and is simply a consequence of the order these integrals appeared in his work.

It is not difficult to see that lattices are subgroups of $\mathbb{C}$ under addition, and that the quotient groups $\mathbb{C}/L$ are isomorphic to the torus group $\mathbb{T}^1 \cong S^1 \times S^1$. Lattices can be visualised in $\mathbb{C}$ as something similar to the following diagram.



Another important property of $t(K,k)$ is that it is meromorphic, meaning that it can be expanded locally as a complex Laurent series everywhere. Jacobi coined the term *elliptic function* to reference all meromorphic, doubly periodic functions.

It cannot be overstated the impact studying elliptic functions has had on mathematics. Whether it be introducing modular forms to number theory or elliptic curves to algebra and geometry, they are the origin of many fascinating topics.

Although the functions $t(K, k)$ are interesting in their own right, studying elliptic functions in general is much more fruitful. Elliptic functions are not uncommon and are relatively easy to construct, so finding them is not an issue. However, classifying them is more difficult. Ideally, for a given lattice L, it would be desirable to have a finite number of elliptic functions over L which generate *all* elliptic functions over L. This scenario is analogous to how smooth, periodic functions can be expanded in terms of sin and cos as Fourier series.

Due to the work of Weierstrass, it is known that two such generators always exist. They are known as the *Weierstrass $\wp$-functions with respect to L.*

For an arbitrary lattice L, the first Weierstrass $\wp$-function is defined as

$$\wp_L(z) := \frac{1}{z^2} + \sum_{\substack{\lambda \in L \\ \lambda \neq 0}} \left(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2} \right).$$

The second Weierstrass $\wp$-function is the derivative of the first divided by a factor[3] of -2,

$$\wp'_L(z) := \sum_{\lambda \in L} \frac{1}{(z-\lambda)^3} \, .$$

The following theorem encapsulates why the Weierstrass $\wp$-functions are essential and why they are regarded as the generators for elliptic functions.

Theorem 2.1. *A complex function $f : \mathbb{C} \longrightarrow \mathbb{C}$ is an elliptic function with respect to L if and only if $f \in \mathbb{C}(\wp_L, \wp'_L)$.*

In the preceding discussion, it was clear that $\wp_L(z)$ and $\wp'_L(z)$ share several properties with $\sin(x)$ and $\cos(x)$. Both pairs are periodic, one over $\mathbb{C}$ and the other over $\mathbb{R}$. Both can be differentiated infinitely often and both can be expanded locally via Laurent series. Also, both pairs act as generators for large classes of important functions.

There are many more similarities they have in common, but the two most consequential will be described in the following sections. Herein, they are referred to as the *Algebraic Law* and the *Addition Law*.

[3] Sometimes this factor is kept, but we exclude it to make life simpler.

2.1 The Algebraic Law

There is no shortage of algebraic relations between sin and cos, but perhaps the most famous among them is $\sin^2(\theta) + \cos^2(\theta) = 1$. This equation can be interpreted as saying the point $(\sin(\theta), \cos(\theta))$ lies on the algebraic curve $x^2 + y^2 = 1$, for any $\theta \in \mathbb{R}$. Provided a unique lattice L is specified, a similar result, outlined below, can be obtained for $\wp_L(z)$ and $\wp'_L(z)$.

First, let $R := \min\{\, |\lambda| \;\big|\; \lambda \in L, \lambda \neq 0\}$. Then, the Laurent expansions of $\wp_L$ and $\wp'_L$ in the disc $|z| < R$ are of the form

$$\wp_L(z) = \frac{1}{z^2} + g(z)$$

and

$$\wp'_L(z) = \frac{1}{z^3} + h(z)$$

where $g(z)$ and $h(z)$ are holomorphic in this disc.

Furthermore, it is straightforward to verify that

$$\wp'_L(z)^2 - \wp_L(z)^3 = \frac{A}{z^2} + i(z),$$

where $i(z)$ is holomorphic in the disc $|z| < R$ and A is a constant depending on L.

These equations imply that $\wp'_L(z)^2 - \wp_L(z)^3 - A\wp_L(z)$ must be holomorphic in a neighbourhood of zero.

From the definition of $\wp_L$ and $\wp'_L$ it is clear that the only possible poles of the function $\wp'_L(z)^2 - \wp_L(z)^3 - A\wp_L(z)$ are the lattice points $\lambda \in L$. However, since $\wp'_L(z)^2 - \wp_L(z)^3 - A\wp_L(z)$ is holomorphic at zero, it must be holomorphic at each $\lambda \in L$ due to the fact it is doubly periodic. Hence it is entire, bounded, and thus, constant by Liouville's theorem.

It follows immediately that

$$\wp'_L(z)^2 = \wp_L(z)^3 + A\wp_L(z) + B$$

for some A,$B \in \mathbb{C}$, both dependent on L.

This result can be rephrased as saying $(\wp_L(z), \wp'_L(z))$ is a solution to the algebraic curve $y^2 = x^3 + Ax + B$ for any $z \in \mathbb{C}$. For reasons that will become apparent later, the discriminant of this cubic i.e. $\Delta := 4A^3 + 27B^2$ will never be zero. When an algebraic curve of the form $y^2 = x^3 + Ax + B$ has a non-zero cubic discriminant, it is known as an *elliptic curve*. These curves are central to modern mathematics and are the primary object of study in this thesis. The Algebraic Law summarises what was discussed above. The converse is also included for completeness.

The Algebraic Law. *For any lattice L and any $z \in \mathbb{C}$, the point $(\wp_L(z), \wp'_L(z))$ will always lie on an elliptic curve $y^2 = x^3 + Ax + B$, where A and B are dependent on L. Furthermore, every elliptic curve corresponds to a unique lattice in $\mathbb{C}$.*

2.2 The Addition Law

Another well known result in trigonometry is the existence of a group homomorphism between $(\mathbb{R}/\mathbb{Z}, +)$ and the circle group S^1 via the map $\theta \mapsto (\sin(\theta), \cos(\theta))$. The fact such a homomorphism exists is a result of the addition formulas:

$$\begin{aligned} \sin(\theta_1 + \theta_2) &= \sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2), \\ \cos(\theta_1 + \theta_2) &= \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2). \end{aligned}$$

Using $\wp_L$ and $\wp'_L$, similar formulas can be constructed for $(\mathbb{C}/L, +)$. These formulas can then be used to induce a group structure on elliptic curves. This is reminiscent of how a group law for the curve $x^2 + y^2 = 1$ can be determined by examining the above equations.

The formulas for $\wp_L$ and $\wp'_L$ come in several cases. Case 1 is the most important as it is the most commonly used and is the basis for the others. The remaining three cases all deal with technicalities of limits.

Case 1. *When* $\wp_L(z_1) \neq \wp_L(z_2)$ *then*

$$\wp_L(z_1 + z_2) = \Big(\frac{\wp'_L(z_2) - \wp'_L(z_1)}{\wp_L(z_2) - \wp_L(z_1)}\Big)^2 - (\wp_L(z_1) + \wp_L(z_2))$$

and

$$\wp'_L(z_1 + z_2) = \frac{1}{2}\Big(\frac{\wp'_L(z_2) - \wp'_L(z_1)}{\wp_L(z_2) - \wp_L(z_1)}\Big)\Big(\wp_L(z_1) - \wp_L(z_1 + z_2)\Big) - \wp'_L(z_1).$$

Case 2. *If* $\wp_L(z)$ *has a pole at* z_1 *and* z_2*, then* $\wp_L(z)$ *and* $\wp'_L(z)$ *have a pole at* $z_1 + z_2$*.*

Case 3. *If* $\wp_L(z_1) = \wp_L(z_2)$ *but* $\wp'_L(z_1) \neq \wp'_L(z_2)$*, then* $\wp_L(z)$ *and* $\wp'_L(z)$ *have a pole at* $z_1 + z_2$*.*

Case 4. *If* $\wp_L(z_1) = \wp_L(z_2)$ *and* $\wp'_L(z_1) = \wp'_L(z_2)$*, then*

$$\wp_L(z_1 + z_2) = \Big(\frac{3\wp_L(z_1)^2 + A}{2\wp'_L(z_1)}\Big)^2 - 2\wp_L(z_1)$$

and

$$\wp'_L(z_1 + z_2) = \Big(\frac{3\wp_L(z_1)^2 + A}{2\wp'_L(z_1)}\Big)(\wp_L(z_1) - \wp'_L(z_1 + z_2)) - \wp'_L(z_1).$$

In the fourth case, A is the same as before (it is the coefficient of x in the elliptic curve associated to the lattice L).

The collection of formulas above is the Addition Law. The proof of these formulas is quite long, so for the sake of cohesion it is excluded. A full proof can be found in [**30**], chapter 9.

The Addition Law describes how a group law can be induced on elliptic curves via addition over $\mathbb{C}$. The group induced on the elliptic curve E is denoted $E(\mathbb{C})$. To ensure the elliptic curves are groups, it is necessary to add a point at infinity (denoted '∞') which acts as the identity element. This process can be defined rigorously using projective geometry[4].

[4]See Appendix

There are several cases to consider when defining the group operation, which is denoted '+'. Let (x_1, y_1) and (x_2, y_2) be points on an elliptic curve $y^2 = x^3 + Ax + B$ with $\Delta \neq 0$, then the group structure is defined as follows.

Case 1. *If* $x_1 \neq x_2$, *let* $m = \frac{y_2 - y_1}{x_2 - x_1}$. *Then* $(x_1, y_1)+(x_2, y_2) = (x_3, y_3)$ *where* $x_3 = m^2 - x_1 - x_2$ *and* $y_3 = m(x_1 - x_3) - y_1$.

Case 2. $(x_1, y_1) + \infty = \infty + (x_1, y_1) = (x_1, y_1)$ *and* $\infty + \infty = \infty$.

Case 3. *If* $x_1 = x_2$ *but* $y_1 \neq y_2$, *then* $y_1 = -y_2$ *and* $(x_1, y_1) + (x_2, y_2) = \infty$.

Case 4. *Suppose* $x_1 = x_2$ *and* $y_1 = y_2$ *and let* $m = \frac{3x_1^2 + A}{2y_1}$ *for* $y_1 \neq 0$. *Then* $(x_1, y_1) + (x_2, y_2) = (x_3, y_3)$ *where* $x_3 = m^2 - 2x_1$ *and* $y_3 = m(x_1 - x_3) - y_1$. *If* $y_1 = 0$, *then* $(x_1, y_1) + (x_2, y_2) = \infty$.

By substituting $x_i = \wp_L(z_i)$, $y_i = \wp'_L(z_i)$ and $x_3 = \wp_L(z_1 + z_2)$, $y_3 = \wp'_L(z_1 + z_2)$ into these formulas, the Addition Law[5] reappears.

Cases 2, 3 and 4 define the identity of the group as ∞ and the inverse of (x, y) as $(x, -y)$. Showing that the group operation '+' is closed and commutative is relatively easy by examination. However, without the use of the Addition Law, associativity requires a great deal of algebraic manipulation to prove and will not be done here.

It is not difficult to extend the definition of an elliptic curve to general fields. This will be the subject of the next section. For simplicity, fields $\mathbb{F}$ with $\mathrm{Char}(\mathbb{F}) = 2, 3$ are excluded in this thesis as issues arise with powers of 2 and powers of 3, though it is still possible to define elliptic curves over these fields. For more, see [**30**], chapter 2.

[5]Observant readers will notice a factor of $\frac{1}{2}$ is missing from Case 1. This does not effect the group structure. It is removed so that the elliptic curve is monic in both variables.

2.3 General Elliptic Curves

In the previous sections, an elliptic curve was defined as an algebraic curve of the form $y^2 = x^3 + Ax + B$ where $A, B \in \mathbb{C}$ and $\Delta = 4A^3 + 27B^2 \neq 0$. An additional point at infinity was added to give a group structure. However, this is just a special case. There is no reason why the definition of an elliptic curve cannot be extended to other fields. The general definition is as follows.

Definition 2.2. *Let $\mathbb{F}$ be a field and $\mathbb{L}$ be a subfield of $\mathbb{F}$. Suppose $A, B \in \mathbb{L}$ and $\Delta := 4A^3 + 27B^2 \neq 0$. Then $E : y^2 = x^3 + Ax + B$ is an **elliptic curve defined over** $\mathbb{L}$. The set of $\mathbb{F}$-rational points on this curve is denoted*

$$E(\mathbb{F}) := \{(x, y) \in \mathbb{F} \times \mathbb{F} \,\big|\, y^2 = x^3 + Ax + B\} \cup \{\infty\}.$$

Two important examples are;

$$E(\mathbb{Q}) := \{(x, y) \in \mathbb{Q} \times \mathbb{Q} \,\big|\, y^2 = x^3 + Ax + B\} \cup \{\infty\}.$$

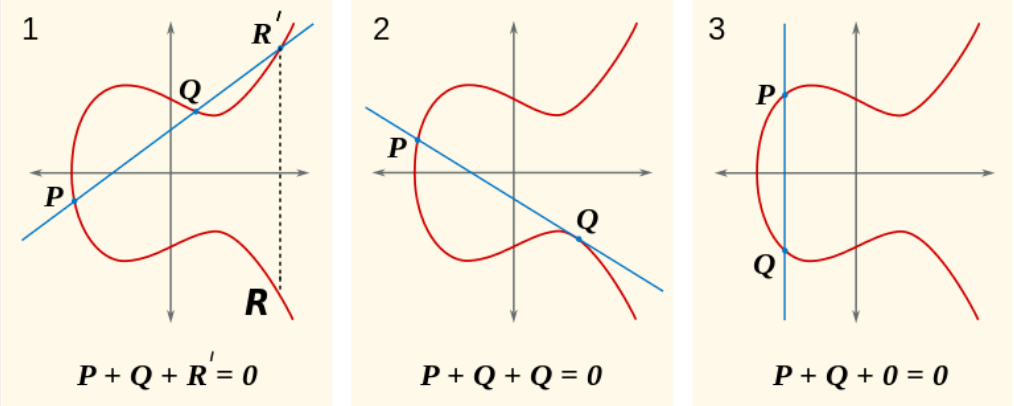
and

$$E(\mathbb{R}) := \{(x, y) \in \mathbb{R} \times \mathbb{R} \,\big|\, y^2 = x^3 + Ax + B\} \cup \{\infty\}.$$

As one would expect, elliptic curves defined over $\mathbb{L}$ also have a group structure. The formulas in Cases 1 to 4 of the previous section defined a group law for elliptic curves over $\mathbb{C}$. However, each formula makes sense and is well-defined over the field $\mathbb{L}$, since it contains the element A and has $\text{Char}(\mathbb{L}) \neq 2, 3$ by assumption. Hence the same algebraic manipulation that proved the group $E(\mathbb{C})$ was closed, abelian and associative also works for $E(\mathbb{L})$. Inverses remain the same i.e. $-(x, y) = (x, -y)$, while ∞ still acts as the identity.

Note that elliptic curves defined over $\mathbb{Q}$ are often referred to as *rational elliptic curves*.

When working over $\mathbb{R}$, it is possible to interpret the group $E(\mathbb{R})$ geometrically. Let $P, Q \in E(\mathbb{R})$. Then $-(P + Q)$ will be the unique point in $E(\mathbb{R})$ that is co-linear with P and Q. The point $P + Q$ will be the reflection of $-(P + Q)$ about the x-axis, since $-(x, y) := (x, -y)$ for all $(x, y) \in E(\mathbb{R})$. This idea is illustrated by the diagram on the next page. The red curve is the set $E(\mathbb{R})$.



In the first figure, the points R and R' denote $P + Q$ and $-(P + Q)$ respectively. The second figure explains how a point, in this case Q is added to itself. The tangent line at Q intersects the elliptic curve at the point P, which denotes $-(Q + Q)$. The reflection of P about the x-axis gives $-P = Q + Q$. In order to be able to define a tangent line at every point on the curve, it is necessary and sufficient that Δ be non-zero. If that were not the case then the curve would contain self-intersections. The third diagram reiterates the fact that the inverse of a point P on an elliptic curve is its reflection about the x-axis, Q. Note that 0 is different notation for ∞.

Now that elliptic curves have been defined for general fields, our next goal will be to understand them. Depending on the field of interest, this can be relatively easy or incredibly hard. In fact, a complete description of the rational case is still unknown. One of the most plausible descriptions to date arose from the work of Birch and Swinnerton-Dyer in their seminal papers [**26**] and [**27**], as we shall see. It is known as the **Birch and Swinnerton-Dyer conjecture**.

2.4 Classifying Elliptic Curve Groups

One of the main goals in group theory is to classify groups up to isomorphism. For elliptic curve groups, this is a highly non-trivial problem and has only been resolved completely in special cases. Two of the most famous examples are:

Theorem 2.2. *Let E be an elliptic curve defined over $\mathbb{C}$. Then $E(\mathbb{C}) \cong S^1 \times S^1$ where S^1 is the circle group.*

Theorem 2.3. *Let E be an elliptic curve defined over $\mathbb{R}$. Then $E(\mathbb{R}) \cong S^1$ if $\Delta < 0$ and $E(\mathbb{R}) \cong \mathbb{Z}_2 \times S^1$ if $\Delta > 0$.*

A proof of the first fact can be found in [**30**], chapter 9. The second follows from general results in Lie theory.

For an elliptic curve E defined over a finite field $\mathbb{F}_{p^n}$, finding the isomorphism class of $E(\mathbb{F}_{p^n})$ is difficult but not impossible. In fact since $\mathbb{F}_{p^n}$ is finite, $E(\mathbb{F}_{p^n})$ is too, so this problem can be resolved in a finite number of steps, by hand or by computer. Thus, the most interesting cases are the elliptic curve groups defined over fields of characteristic zero. Since $\mathbb{Q}$ is the smallest[6] such field, classifying $E(\mathbb{Q})$ is essential.

In 1901, Henri Poincaré proposed that any point on a rational elliptic curve $E(\mathbb{Q})$ should be expressible as a sum of finitely many infinite order points on $E(\mathbb{Q})$ and finitely many finite order points on $E(\mathbb{Q})$ (addition in this case being the group operation for elliptic curves). The minimum number of infinite order points required to do this is referred to as the **rank** of the curve. The points of finite order are referred to as **torsion points**. A modern formulation of his conjecture is

Conjecture. *Let E be an elliptic curve defined over $\mathbb{Q}$, then $E(\mathbb{Q})$ is a finitely generated group.*

[6] In the sense that every field of characteristic zero contains a subfield isomorphic to $\mathbb{Q}$. It is the only field with this property.

Since $E(\mathbb{Q})$ is always commutative, the classification of finitely generated abelian groups can be used to restate the conjecture.

Conjecture. *Let E be an elliptic curve defined over $\mathbb{Q}$, then $E(\mathbb{Q})$ is isomorphic to a finite direct product of cyclic groups i.e.*

$$E(\mathbb{Q}) \cong \mathbb{Z}^r \times \prod_{i=1}^{N} \mathbb{Z}_{q_i}$$

where r is the rank of E, N depends on E and each q_i is a prime power. The product of the finite cyclic groups forms a subgroup denoted T which is sometimes referred to as the ***torsion subgroup****.*

A corollary of the above conjecture is that the rank of an elliptic curve will always be finite.

Two decades after it was posed, Poincaré's conjecture was finally proved by Louis J. Mordell in 1922, see [**19**]. The result is named after Mordell in honour of his proof. Since its discovery, Mordell's theorem has been significantly generalised and has gone on to become a foundational result in number theory and algebraic geometry.

Mordell's Theorem. *Let E be an elliptic curve defined over $\mathbb{Q}$, then*

$$E(\mathbb{Q}) \cong \mathbb{Z}^r \times T$$

for some $r \in \mathbb{N} \cup \{0\}$ and some finite group T.

Proving Mordell's theorem is rather difficult and would take too long to complete in this thesis. Instead, the basic ideas and strategy of the proof are described. Complete proofs can be found in [**19**], [**23**] and [**30**].

Firstly, the introduction of two definitions is required.

Definition 2.3. *Let $\left(\frac{a}{b}, \frac{c}{d}\right)$ be a point on $E(\mathbb{Q})$, written so that a and b are coprime. The* ***multiplicative height*** *$H : E(\mathbb{Q}) \longrightarrow \mathbb{R}^+$ is defined as $H\left(\frac{a}{b}, \frac{c}{d}\right) := \max\{|a|, |b|\}$. Furthermore, $H(\infty) = 1$ by assumption.*

Definition 2.4. *The* ***logarithmic height*** *$h : E(\mathbb{Q}) \longrightarrow \mathbb{R}^+$ is defined as $h(P) := \log(H(P))$. Furthermore, $h(\infty) = 0$ by assumption.*

These functions play an important role as they quantify how large the numerators and denominators of rational points on elliptic curve are. Studying how the values of the height functions change under the group law on $E(\mathbb{Q})$ is at the heart of Mordell's proof.

The following properties of the logarithmic height are essential for the proof. The corresponding properties for the mulitplicative height can be obtained by exponentiating.

Property 1. *Let P and Q be arbitrary points on $E(\mathbb{Q})$. Then there exists a constant C_1 depending on E and one of the points (say Q) such that*

$$h(P+Q) \leq 2h(P) + C_1.$$

Property 2. *Let P be an arbitrary point on $E(\mathbb{Q})$. Then there exists a constant C_2 depending on E such that*

$$h(2P) \geq 4h(P) - C_2.$$

Property 3. *Let C_3 be any positive constant. Then there are only finitely many $P \in E(\mathbb{Q})$ such that $h(P) \leq C_3$.*

The final result needed for the proof is group theoretic in nature. It is what allows us to translate the above properties into information about group structure.

The Weak Mordell Theorem. *$E(\mathbb{Q})/2E(\mathbb{Q})$ is a finite group.*

Proving this result is the most difficult part of Mordell's theorem, which is why it is known as the weak Mordell theorem. A proof is provided below in the case where $E : y^2 = x^3 + Ax + B$ only has rational roots i.e. $x^3 + Ax + B = 0$ only has rational solutions. The proof is essentially the same as [**30**], chapter 8. It is accomplished through several lemmas, which will be delineated now.

The first lemma provides a canonical form for these elliptic curves.

Lemma 2.1. *Suppose $E : y^2 = x^3 + Ax + B$ is defined over $\mathbb{Q}$ and that the polynomial $x^3 + Ax + B = 0$ only has rational roots. Then, there exists an elliptic curve of the form $E' : y^2 = (x - n_1)(x - n_2)(x - n_3)$ with $n_i \in \mathbb{Z}$ such that $E(\mathbb{Q}) \cong E'(\mathbb{Q})$.*

Proof. Since E only has rational roots, it can be written as

$$y^2 = \left(x - \frac{a_1}{b_1}\right)\left(x - \frac{a_2}{b_2}\right)\left(x - \frac{a_3}{b_3}\right)$$

where $\frac{a_i}{b_i} \in \mathbb{Q}$.

By substitution, it is easily verified that for each $n \in \mathbb{N}$, the rational map $(x, y) \longrightarrow \left(\frac{x}{n^2}, \frac{y}{n^3}\right)$ is an isomorphism between $E(\mathbb{Q})$ and $E_n(\mathbb{Q})$, where E_n is defined as $y^2 = x^3 + An^4x + Bn^6$.

Thus, if $n = b_1b_2b_3$, $E(\mathbb{Q})$ will be isomorphic to the group of rational points on

$$\left(\frac{y}{(b_1b_2b_3)^3}\right)^2 = \left(\frac{x}{(b_1b_2b_3)^2} - \frac{a_1}{b_1}\right)\left(\frac{x}{(b_1b_2b_3)^2} - \frac{a_2}{b_2}\right)\left(\frac{x}{(b_1b_2b_3)^2} - \frac{a_3}{b_3}\right).$$

Multiplying both sides by $(b_1b_2b_3)^6$ removes all denominators but does not change the solutions to the above equation. Thus, $E(\mathbb{Q})$ will be isomorphic to the group of rational points on an elliptic curve of the form

$$y^2 = (x - n_1)(x - n_2)(x - n_3) \tag{2.1}$$

where the n_i are integers. Note that since the discriminant of an elliptic curve is never zero, these three integers are all distinct. □

Therefore, it can be assumed without loss of generality that $E : y^2 = x^3 + Ax + B$ can always be rewritten in the form (2.1). This assumption will be made for the remainder of this chapter.

The next step is to consider an important map[7] defined on $E(\mathbb{Q})$. This map can be constructed through Galois cohomology, a topic that will be discussed later. The symbol $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ denotes the multiplicative group of rationals modulo squares. In what follows $\equiv$ will denote equality in this quotient group. For an example of how this group works, consider $rs^2 \in \mathbb{Q}^*$, where $r, s \in \mathbb{Q}^*$. Then $rs^2 \equiv r$ in the group $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ because square terms are necessarily trivial.

The map of interest is

$$\psi : E(\mathbb{Q}) \longrightarrow \mathbb{Q}^*/(\mathbb{Q}^*)^2 \times \mathbb{Q}^*/(\mathbb{Q}^*)^2 \times \mathbb{Q}^*/(\mathbb{Q}^*)^2$$

defined by the following cases

$$\begin{aligned}
&(x, y) \mapsto (x - n_1, x - n_2, x - n_3) \text{ if } y \neq 0,\\
&(n_1, 0) \mapsto ((n_1 - n_2)(n_1 - n_3), n_1 - n_2, n_1 - n_3),\\
&(n_2, 0) \mapsto (n_2 - n_1, (n_2 - n_1)(n_2 - n_3), n_2 - n_3),\\
&(n_3, 0) \mapsto (n_3 - n_1, n_3 - n_2, (n_3 - n_1)(n_3 - n_2)),\\
&\infty \mapsto (1, 1, 1).
\end{aligned}$$

[7]This map comes from the classical process of 2-descent, an account of which is provided in [**6**] and [**30**].

The group operation on $\mathbb{Q}^*/(\mathbb{Q}^*)^2 \times \mathbb{Q}^*/(\mathbb{Q}^*)^2 \times \mathbb{Q}^*/(\mathbb{Q}^*)^2$ is component-wise multiplication inherited from $\mathbb{Q}^*/(\mathbb{Q}^*)^2$. The next lemma describes why this map is so important.

Lemma 2.2. *The map ψ is a homomorphism.*

Proof. Let $(x_1, y_1), (x_2, y_2)$ and (x_3, y_3) be three co-linear points on $E(\mathbb{Q})$. This is equivalent to saying $(x_1, y_1) + (x_2, y_2) = (x_3, -y_3)$. Assume for the time being that $y_i \neq 0$ for $i \in \{1, 2, 3\}$.

Let $y = mx + c$ be the line passing through the three points. This line passes through two distinct rational points so m and c are themselves rational. Then, by definition, the equation $(x - n_1)(x - n_2)(x - n_3) - (mx + c)^2 = 0$ has roots at $x = x_1, x_2$ and x_3, thus,

$$(x - n_1)(x - n_2)(x - n_3) - (mx + c)^2 = (x - x_1)(x - x_2)(x - x_3).$$

Evaluating this polynomial at n_1, n_2 and n_3, it becomes clear that for each $i \in \{1, 2, 3\}$, one has

$$(x_1 - n_i)(x_2 - n_i)(x_3 - n_i) \equiv 1 \text{ in } \mathbb{Q}^*/(\mathbb{Q}^*)^2. \tag{2.2}$$

This is equivalent to saying

$$\psi(x_1, y_1)\psi(x_2, y_2)\psi(x_3, y_3) \equiv (1, 1, 1) \text{ in } \mathbb{Q}^*/(\mathbb{Q}^*)^2 \times \mathbb{Q}^*/(\mathbb{Q}^*)^2 \times \mathbb{Q}^*/(\mathbb{Q}^*)^2. \tag{2.3}$$

Next, let $a \in \mathbb{Q}^*$. Using the fact that $a^2 \equiv 1$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ it follows immediately that for any $a_1, a_2, a_3 \in \mathbb{Q}^*$

$$(a_1, a_2, a_3)^2 \equiv (a_1^2, a_2^2, a_3^2) \equiv (1, 1, 1) \tag{2.4}$$

in $\mathbb{Q}^*/(\mathbb{Q}^*)^2 \times \mathbb{Q}^*/(\mathbb{Q}^*)^2 \times \mathbb{Q}^*/(\mathbb{Q}^*)^2$.

Multiplying both sides of (2.3) by $\psi(x_3, y_3)$, (2.4) implies

$$\psi(x_1, y_1)\psi(x_2, y_2) \equiv \psi(x_3, y_3).$$

Furthermore, since $\psi(x, y)$ does not depend on the sign of y, it follows that $\psi(x, y) \equiv \psi(x, -y)$. This implies that

$$\psi(x_1, y_1)\psi(x_2, y_2) \equiv \psi(x_3, -y_3) \equiv \psi((x_1, y_1) + (x_2, y_2))$$

by definition of $(x_3, -y_3)$. Hence ψ is a homomorphism when $y_i \neq 0$.

The only non-trivial case left to prove is when exactly one point (x_i, y_i), $i \in \{1, 2, 3\}$, satisfies $y_i = 0$. The cases where there are two or more such points are easily verified. Suppose then, without loss of generality, that $(x_1, y_1) = (n_1, 0)$. By the above discussion, it suffices to prove

$$\psi(x_1, y_2)\psi(x_2, y_2)\psi(x_3, y_3) \equiv (1, 1, 1)$$

for ψ to be a homomorphism.

By definition of ψ it is easy to verify that (2.2) continues to hold for $i \in \{2, 3\}$ i.e.

$$(x_1 - n_i)(x_2 - n_i)(x_3 - n_i) \equiv 1 \text{ in } \mathbb{Q}^*/(\mathbb{Q}^*)^2 \text{ for } i \in \{2, 3\}. \tag{2.5}$$

Thus, the proof relies on showing that

$$(n_1 - n_2)(n_1 - n_3)(x_2 - n_1)(x_3 - n_1) \equiv 1. \tag{2.6}$$

Since $(x_2, y_2), (x_3, y_3) \in E(\mathbb{Q})$, one has $y_2^2 = (x_2 - n_1)(x_2 - n_2)(x_2 - n_3)$ and $y_3^2 = (x_3 - n_1)(x_3 - n_2)(x_3 - n_3)$, hence

$$\begin{aligned}(x_2 - n_1)(x_2 - n_2)(x_2 - n_3) &\equiv 1\\ (x_3 - n_1)(x_3 - n_2)(x_3 - n_3) &\equiv 1.\end{aligned}$$

Multiplying both sides of the top equation by $(x_2 - n_1)$ and both sides of the bottom equation by $(x_3 - n_1)$, it follows from (2.4) that

$$\begin{aligned}(x_2 - n_1) &\equiv (x_2 - n_2)(x_2 - n_3),\\ (x_3 - n_1) &\equiv (x_3 - n_2)(x_3 - n_3).\end{aligned}$$

Using (2.4) and (2.5) this can be rewritten as

$$\begin{aligned}(x_2 - n_1) &\equiv (x_1 - n_2)(x_3 - n_2)(x_2 - n_3),\\ (x_3 - n_1) &\equiv (x_2 - n_2)(x_1 - n_2)(x_3 - n_3).\end{aligned}$$

Then, taking the product of $x_2 - n_1$ and $x_3 - n_1$, and using (2.5) again gives

$$(x_2 - n_1)(x_3 - n_1) \equiv (x_1 - n_3)(x_1 - n_2).$$

Finally, since $x_1 = n_1$, it follows that

$$(x_2 - n_1)(x_3 - n_1) \equiv (n_1 - n_2)(n_1 - n_3).$$

Multiplying both sides by $(n_1 - n_2)(n_1 - n_3)$ gives (2.6). Hence ψ is a homomorphism.

□

The next lemma provides a complete description of the kernel of ψ.

Lemma 2.3. *The kernel of ψ is the subgroup $2E(\mathbb{Q})$.*

Proof. Since $2E(\mathbb{Q})$ is trivially contained in $\mathrm{Ker}(\psi)$, it suffices to prove $\mathrm{Ker}(\psi) \subseteq 2E(\mathbb{Q})$. This is equivalent to saying that $(x, y) \in 2E(\mathbb{Q})$ if $\psi(x, y) \equiv (1, 1, 1)$.

Suppose then $(x, y) \in E(\mathbb{Q})$ and $\psi(x, y) \equiv (1, 1, 1)$. This implies, for each $i \in \{1, 2, 3\}$, that there exists $r_i \in \mathbb{Q}$ such that

$$(x - n_i) = r_i^2.$$

Now, define $p(t) = at^2 + bt + c$ to be the unique quadratic polynomial such that $p(n_i) = r_i$. The polynomial $x - t - p(t)^2$ will then have roots at each n_i as a result. Equivalently,

$$x - t - p(t)^2 = (t - n_1)(t - n_2)(t - n_3)f(t)$$

for some $f(t) \in \mathbb{Q}[t]$. By applying lemma 2.1 this can be rewritten as

$$x - t - p(t)^2 = \left(t^3 + At + B\right) f(t).$$

By expanding $p(t)^2$ and using the fact

$$\begin{aligned} t^3 &= -At - B + (t^3 + At + B), \\ t^4 &= -At^2 - Bt + (t^3 + At + B). \end{aligned}$$

it can be shown there exists $q(t) \in \mathbb{Z}[t]$ such that for all $t \in \mathbb{C}$,

$$x - t = (b^2 + 2ac - Aa^2)t^2 + (2bc - 2Aab - Ba^2)t + c^2 - 2Bab + (t^3 + At + B)q(t).$$

Since both $x - t$ and $(b^2 + 2ac - Aa^2)t^2 + (2bc - 2Aab - Ba^2)t + c^2 - 2Bab$ have degrees less than three and differ from each other by a multiple of a cubic polynomial, they must be equal. In other words

$$\begin{aligned} &b^2 + 2ac - Aa^2 = 0, \\ &2bc - 2Aab - Ba^2 = -1, \\ &c^2 - 2Bab = x. \end{aligned}$$

Now, suppose for the sake of contradiction that $a = 0$. By the first equation this would imply $b = 0$, which would mean $p(t)^2$ is a constant polynomial and $n_1 = n_2 = n_3$, which contradicts the fact $\Delta \neq 0$. So a is non-zero.

If the first and second equations are multiplied by $\frac{b}{a^3}$ and $\frac{-1}{a^2}$ respectively, they become

$$\begin{aligned}\left(\frac{b}{a}\right)^3 + \frac{2bc}{a^2} - A\left(\frac{b}{a}\right) &= 0,\\ \frac{-2bc}{a^2} + 2A\left(\frac{b}{a}\right) + B &= \frac{1}{a^2}.\end{aligned}$$

Adding both these equations together gives

$$\left(\frac{\pm 1}{a}\right)^2 = \left(\frac{b}{a}\right)^3 + A\left(\frac{b}{a}\right) + B.$$

Therefore, $\left(\frac{b}{a}, \frac{1}{a}\right)$ and $\left(\frac{b}{a}, \frac{-1}{a}\right)$ are elements of $E(\mathbb{Q})$.

To complete the proof, it suffices to show that either $2\left(\frac{b}{a}, \frac{1}{a}\right) = (x, y)$ or that $2\left(\frac{b}{a}, \frac{-1}{a}\right) = (x, y)$.

For an arbitrary point $(x, y) \in E(\mathbb{Q})$, Case 4 of the Addition Law on page 16, implies the following

$$2(\tilde{x}, \pm\tilde{y}) = (x, y) \iff x = \frac{\tilde{x}^4 - 2A\tilde{x}^2 - 8B\tilde{x} + A^2}{4\tilde{y}^2}.$$

With this in mind, solve the equation $b^2+2ac-Aa^2 = 0$ for c and substitute into $c^2-2Bab = x$ to get

$$x = \frac{A^2a^4 - 2Aa^2b^2 + b^4 - 8Ba^3b}{4a^2}.$$

Dividing by a^4 in the numerator and denominator completes the proof. □

The final lemma ties everything together.

Lemma 2.4. *The image of ψ is finite.*

Proof. Every rational number can be written uniquely as the product of a square rational and a squarefree integer. Thus, it suffices to prove that there are only finitely many squarefree integers a, b, c such that

$$\begin{aligned} x - n_1 &\equiv a \\ x - n_2 &\equiv b \\ x - n_3 &\equiv c \end{aligned}$$

where $y^2 = (x - n_1)(x - n_2)(x - n_3)$.

To start, let p be prime and define the p-adic valuation of an integer m as

$$\nu_p(m) := \begin{cases} \max\{k \in \mathbb{N} \cup \{0\} \mid p^k \text{ divides } m\} & \text{if } m \neq 0, \\ \infty & \text{if } m = 0. \end{cases}$$

This can be extended to the rationals via the formula $\nu_p\left(\frac{m}{n}\right) = \nu_p(m) - \nu_p(n)$.

Now suppose p is a prime dividing a and x is rational. Since a is squarefree, this means $\nu_p(x-n_1)$ must be odd. Furthermore, if $\nu_p(x-n_1)$ is negative, it is not difficult to show that $\nu_p(x - n_1) = \nu_p(x - n_2) = \nu_p(x - n_3)$, since adding integers to reduced fractions will not change the denominator. However, then $\nu_p((x - n_1)(x - n_2)(x - n_3)) = 3\nu_p(x - n_1)$ is odd, contradicting the fact $\nu_p((x - n_1)(x - n_2)(x - n_3)) = \nu_p(y^2) = 2\nu_p(y)$ is even. Therefore, $\nu_p(x-n_1)$ must be positive i.e. $p \mid x-n_1$. A corollary of this is that $\nu_p(x-n_2)$ and $\nu_p(x-n_3)$ must be non-negative too.

Next, suppose for the sake of contradiction that $p \nmid (n_1 - n_2)(n_2 - n_3)(n_3 - n_1)$. Then $p \nmid (x - n_2)$ and $p \nmid (x - n_3)$ since $x - n_i = x - n_1 + n_1 - n_i$. Thus, $\nu_p(x - n_2) = \nu_p(x - n_3) = 0$ and

$$\nu_p(y^2) = \nu_p((x - n_1)(x - n_2)(x - n_3)) = \nu_p(x - n_1)$$

which is a contradiction as $\nu_p(x - n_1)$ is odd.

Therefore, every prime that divides a must also divide $(n_1 - n_2)(n_2 - n_3)(n_3 - n_1)$. However, because $(n_1 - n_2)(n_2 - n_3)(n_3 - n_1)$ is constant and a is squarefree, there are only finitely many choices for such an a among the integers. The same idea works for b and c, proving the image of ψ is finite.

□

With these four lemmas in mind, the first isomorphism theorem yields

$$E(\mathbb{Q})/2E(\mathbb{Q}) \cong \mathrm{Im}(\psi).$$

Thus, $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite and the special case of the weak Mordell theorem has now been proved.

To complete this section, a proof of Mordell's theorem using properties 1-3 of the height functions and the weak Mordell theorem is provided.

Mordell's Theorem. *Let E be an elliptic curve defined over $\mathbb{Q}$, then*

$$E(\mathbb{Q}) \cong \mathbb{Z}^r \times T$$

for some $r \in \mathbb{N} \cup \{0\}$ and some finite group T.

Proof. Let $P \in E(\mathbb{Q})$. Since $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite, there are only finitely many coset representatives of $2E(\mathbb{Q})$ in $E(\mathbb{Q})$. The set of representatives is denoted R_E. Thus, P can be written as

$$P = R_1 + 2P_1$$

where $R_1 \in R_E$ and P_1 is defined as a solution to this equation. Furthermore P_1 can be written as

$$P_1 = R_2 + 2P_2$$

where $R_2 \in R_E$, $P_2 \in E(\mathbb{Q})$. These two equations can be combined into

$$P = R_1 + 2R_2 + 4P_2.$$

Continuing this process n times gives

$$P_n = R_{n+1} + 2P_{n+1},$$
$$P = R_1 + 2R_2 + \cdots + 2^{n-1}R_n + 2^n P_n.$$

If it can be shown for large enough n that P_n must be contained in some finite set, then there can only be finitely many choices for P_n and hence only finitely many generators. The theorem would then follow. The remainder of the proof is dedicated to showing this.

Property 1 says

$$h(Q - R_i) \leq 2h(Q) + C_i \text{ for all } Q \in E(\mathbb{Q})$$

where C_i denotes a constant depending on $R_i \in R_E$. Let C be the maximum among these C_i. Such a maximum exists since R_E is finite.

Now use property 2 to get

$$4h(P_i) \leq h(2P_i) + \tilde{C} = h(P_{i-1} - R_i) + \tilde{C} \leq 2h(P_{i-1}) + C + \tilde{C}$$

where $\tilde{C}$ only depends on E.

This implies

$$h(P_i) \leq \frac{3}{4}h(P_{i-1}) - \frac{1}{4}(h(P_{i-1}) - C - \tilde{C}).$$

If $h(P_{i-1}) \leq C + \tilde{C}$ for some $i \in \mathbb{N}$, then we are finished due to property 3. Suppose then $h(P_{i-1}) > C + \tilde{C}$ for all $i \in \mathbb{N}$. This means $h(P_i) < \frac{3}{4}h(P_{i-1})$, so that therefore $h(P_i)$ is a strictly decreasing sequence and eventually $h(P_n) \leq C + \tilde{C}$ for large enough n, a contradiction. Hence there will always exist $n \in \mathbb{N}$ such that $h(P_n) \leq C + \tilde{C}$. By property 3, such P_n must be finite in number and the proof is complete. □

Chapter 3

The Tate-Shafarevich Group

The goal of this chapter is to introduce the *Tate-Shafarevich group*, a fundamental object in algebraic geometry. It will prove useful later when discussing isomorphism classes and the Birch and Swinnerton-Dyer conjecture. With this in mind, it will be necessary to introduce several new concepts.

In the previous section, we concluded that one of the main problems about elliptic curves is classifying the group $E(\mathbb{Q})$ up to isomorphism. The first major step in resolving this problem was Mordell's theorem. As a result of Mordell's theorem, the classification of $E(\mathbb{Q})$ reduces to determining the rank and torsion subgroup of E. This is easier said than done, though major progress has been made in both cases. Indeed, the torsion case was essentially solved by Mazur, [**18**], in 1978, when he proved a long-standing conjecture of Levi:

Mazur's Theorem. *Suppose E is an elliptic curve defined over $\mathbb{Q}$ and T is the torsion group of $E(\mathbb{Q})$, then T must satisfy one of the following:*

$$\begin{aligned} &T \cong \mathbb{Z}_n,\ 1 \leq n \leq 10, \\ &T \cong \mathbb{Z}_{12}, \\ &T \cong \mathbb{Z}_2 \times \mathbb{Z}_{2n},\ 1 \leq n \leq 4. \end{aligned}$$

Furthermore, every group in this list occurs as the torsion group of infinitely many elliptic curves.

The last part guarantees this theorem is necessary and sufficient for classifying torsion groups, meaning it cannot be improved further without altering the hypotheses of the theorem.

Therefore, the most interesting case is determining, or at least bounding the rank of E. The best approach to this problem is to focus on the quotient groups $E(\mathbb{Q})/nE(\mathbb{Q})$ rather than $E(\mathbb{Q})$ itself.

Special attention is given to the case $n = 2$ due to the fact $E(\mathbb{Q})/2E(\mathbb{Q})$ usually has the lowest order among these quotient groups. Moreover, there is a simple but useful relationship between the order of this group and the rank of E.

Proposition. *Let $E : y^2 = x^3 + Ax + B$ be an elliptic curve defined over $\mathbb{Q}$ and ρ be the number of rational solutions to $x^3 + Ax + B = 0$. Suppose that the rank of E is r. Then,*

$$|E(\mathbb{Q})/2E(\mathbb{Q})| = \begin{cases} 2^r & \text{if } \rho = 0 \\ 2^{r+1} & \text{if } \rho = 1 \\ 2^{r+2} & \text{if } \rho = 3 \end{cases}$$

Note that $\rho \neq 2$ since otherwise the sum of the roots would be irrational and non-zero, contradicting the definition of E.

This proposition is a straightforward corollary of Mordell's theorem.

Hence the problem of bounding the rank of an elliptic curve can be reduced to bounding the order of $E(\mathbb{Q})/2E(\mathbb{Q})$, or more generally $E(\mathbb{Q})/nE(\mathbb{Q})$. The most efficient way of solving this problem involves first ascertaining the isomorphism classes of these quotient groups.

For example, in the last section the isomorphism

$$E(\mathbb{Q})/2E(\mathbb{Q}) \cong \mathrm{Im}(\psi)$$

was established and provided essential information about the group $E(\mathbb{Q})/2E(\mathbb{Q})$. The proof of this fact implicitly used Galois cohomology, a useful framework for studying elliptic curves. The fundamentals of Galois cohomology are outlined in the next two sections with the intention of generalising the above result and introducing the Tate-Shafarevich group. To maintain a coherent exposition, several results in these sections will be stated without proof.

3.1 Galois Cohomology

Group cohomology is the application of cohomological principles from algebraic topology to the study of groups. Galois cohomology has the same premise but specialises in Galois groups. This theory can be developed independently of topological spaces by considering group modules instead.

To begin, several terms and definitions are introduced. The first one is particularly important as it encapsulates what it means for two rational algebraic curves to be equivalent in algebraic geometry.

Definition 3.1. *Let D_1 and D_2 be two algebraic curves defined over $\mathbb{Q}$ and let $\mathbb{F}$ be a subfield of $\overline{\mathbb{Q}}$. Then D_1 and D_2 are said to be **birationally equivalent over** $\mathbf{F}$ if there exists a rational bijection $\Phi : D_1 \longrightarrow D_2$ defined over $\mathbb{F}$ whose inverse is also rational i.e.*

$$\Phi(x, y) = (R_1(x, y), R_2(x, y))$$
$$\Phi^{-1}(x, y) = (R_3(x, y), R_4(x, y))$$

*for some $R_i(x, y) \in \mathbb{F}(x, y)$. The map Φ is known as a **birational map**.*

In the case where $D_1 = D_2$, the map Φ is commonly referred to as a *birational automorphism*. Non-trivial automorphisms on a curve can be regarded as symmetries in its structure. Such automorphisms can be constructed via group actions. The group primarily used in this process is the **absolute Galois group** $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, which shall be denoted Γ.

Suppose $\sigma \in \Gamma$ and $\Phi : D_1 \longrightarrow D_2$ is birational (assume it is defined over $\overline{\mathbb{Q}}$ unless otherwise stated). Define the function $\sigma\Phi : D_1 \longrightarrow D_2$ to be the map obtained by letting σ

act on the coefficients of Φ. This map is also birational, which means the composite map $(\sigma\Phi)\Phi^{-1} : D_2 \longrightarrow D_2$ will be a birational automorphism of the curve D_2.

Of course, σ may act trivially on Φ, making $(\sigma\Phi)\Phi^{-1}$ the identity function. However, provided that Φ is defined over some proper field extension of $\mathbb{Q}$, there will always exist some $\sigma \in \Gamma$ which acts non-trivially.

These automorphisms are particularly easy to characterise on elliptic curves due to the underlying group structure. In fact, if E is an elliptic curve defined over $\mathbb{Q}$ then automorphisms of $E(\overline{\mathbb{Q}})$ can be written as

$$(\sigma\Phi)\Phi^{-1} : E(\overline{\mathbb{Q}}) \longrightarrow E(\overline{\mathbb{Q}}) \, , \; P \mapsto P + \Lambda_\Phi(\sigma) \tag{3.1}$$

for some map $\Lambda_\Phi : \Gamma \longrightarrow E(\overline{\mathbb{Q}})$ dependent on the choice of Φ.

Thus, automorphisms of $E(\overline{\mathbb{Q}})$ are merely translations by group elements. A proof of this fact can be found in [**6**], chapter 20.

To understand these automorphisms, it is necessary to focus on the family of functions $\{\Lambda_\Phi\}$. They are examples of *cocycles*, an unilluminating term which has its roots in topology. Cocycles are noteworthy for their functional identity, which in this case is,

$$\Lambda_\Phi(\tau\sigma) = \tau\Lambda_\Phi(\sigma) + \Lambda_\Phi(\tau) \text{ for all } \sigma, \tau \in \Gamma. \tag{3.2}$$

Here $\tau \in \Gamma$ acts on points in $E(\overline{\mathbb{Q}})$ componentwise i.e. $\tau(x, y) = (\tau x, \tau y)$. This action makes $E(\overline{\mathbb{Q}})$ a Γ-module. The identity (3.2) follows from composing the aforementioned automorphisms in (3.1).

Note too that if Γ acts trivially on the image of Λ_Φ, the cocycle identity (3.2) reduces to

$$\Lambda_\Phi(\tau\sigma) = \Lambda_\Phi(\sigma) + \Lambda_\Phi(\tau),$$

making $\Lambda_\Phi : \Gamma \longrightarrow E(\overline{\mathbb{Q}})$ a homomorphism of groups. For this reason, cocycles are often referred to as *twisted homomorphisms* as well.

An important property that cocycles share with homomorphisms is that they form groups. In fact, given any abelian G-module $(R, +)$, the set

$$Z(G, R) := \{f : G \longrightarrow R \mid f(gh) = gf(h) + f(g) \text{ for all } g, h \in G\}$$

will inherit a group structure from $(R, +)$.

In order to translate what has been discussed so far into information about $E(\mathbb{Q})/nE(\mathbb{Q})$, it is necessary to introduce the following.

Definition 3.2. *The **Zeroth cohomology group** of an abelian G-module $(R, +)$, is defined to be the subgroup of R fixed trivially by G,*

$$H^0(G, R) := \{r \in R \mid gr = r \text{ for all } g \in G\}.$$

This definition is introduced as a means of characterising $E(\mathbb{Q})$ in terms of $E(\overline{\mathbb{Q}})$ and Γ as

$$E(\mathbb{Q}) = H^0(\Gamma, E(\overline{\mathbb{Q}})). \tag{3.3}$$

The advantage of this identity is that it allows $E(\mathbb{Q})$ to be studied via $E(\overline{\mathbb{Q}})$, a group much easier to work with. It is also what allows one to study elliptic curves through a cohomological viewpoint.

Now, the task at hand is to establish a relationship between $H^0(G, R)$ and $Z(G, R)$.

To this end, consider the functions $\delta_r : G \longrightarrow R$ defined by $\delta_r(g) = gr - r$, where as before, $(R, +)$ is an abelian G-module, $r \in R$ and $g \in G$.

This function can be used as a means of 'detecting' elements of $H^0(G, R)$, in the sense that if δ_r is the identity for all g, then clearly $r \in H^0(G, R)$. The heuristic view is *the more often δ_r is trivial, the more likely r is in $H^0(G, R)$.*

Another important property of the functions δ_r, which is proved below, is that they are cocycles. They are also known as *coboundaries*. In a certain sense, coboundaries are the 'trivial' cocycles since they arise naturally and are easily constructed. On the other hand, coboundaries prevent G from acting trivially on $Z(G, R)$. This makes the classification of cocycles and birational automorphisms much more difficult, so removing coboundaries is a priority.

It shall now be proved that the set of all coboundaries, denoted $B(G,R)$, forms a subgroup of $Z(G,R)$.

Proof. By definition,

$$B(G,R) := \{\delta_r : G \longrightarrow R \mid r \in R,\ \delta_r(g) = gr - r\}$$

Now suppose $\delta_r \in B(G,R)$, then for all $g, h \in G$,

$$\delta_r(gh) = ghr - r.$$

However, one also has

$$ghr - r = (ghr - gr) + (gr - r) = g(\delta_r(h)) + \delta_r(g).$$

So $B(G,R)$ is indeed a subset of $Z(G,R)$. To prove $B(G,R)$ is a subgroup it suffices to prove closure and the existence of inverses.

Suppose then $\delta_r, \delta_s \in B(G,R)$. We need to show $\delta_r + \delta_s \in B(G,R)$. However, this is immediate from the fact R is closed under + and

$$(\delta_r + \delta_s)(g) := \delta_r(g) + \delta_s(g) = gr - r + gs - s = g(r+s) - (r+s) = \delta_{r+s}(g) \text{ for all } g \in G.$$

Therefore, $\delta_r + \delta_s = \delta_{r+s}$. This identity also makes it clear that δ_{-r} is the inverse of δ_r. □

The subgroup $B(G,R)$ is automatically normal because $Z(G,R)$ inherits the commutativity of R. Since we want to remove coboundaries from future considerations, it is natural to focus on the corresponding quotient group detailed below.

Definition 3.3. *The **First cohomology group** of an abelian G-module $(R,+)$, is defined to be the quotient group*

$$H^1(G,R) := Z(G,R)/B(G,R).$$

In particular,

$$H^1(\Gamma, E(\overline{\mathbb{Q}})) := Z(\Gamma, E(\overline{\mathbb{Q}}))/B(\Gamma, E(\overline{\mathbb{Q}})).$$

By factoring out the coboundary subgroup, the difficulty mentioned earlier is removed. In other words, the following lemma holds.

Lemma 3.1. *The group G acts trivially on $H^1(G, R)$.*

A proof of this lemma can be found in [**6**], chapter 21.

What makes cohomology groups particularly consequential is that they allow short exact sequences to be extended. This is useful in so far as one can use this fact to re-characterise $E(\mathbb{Q})/nE(\mathbb{Q})$ from a cohomological viewpoint.

Consider the short exact sequence

$$0 \to E(\overline{\mathbb{Q}})[n] \xrightarrow{\iota} E(\overline{\mathbb{Q}}) \xrightarrow{n} E(\overline{\mathbb{Q}}) \to 0,$$

where $E(\overline{\mathbb{Q}})[n]$ denotes the elements of $E(\overline{\mathbb{Q}})$ with order dividing n. Also ι and n denote the inclusion and multiplication-by-n homomorphisms respectively. Surjectivity of the homomorphism n is proven in [**30**], chapter 2.

This short exact sequence can be used to induce the long exact sequence

$$\begin{aligned} 0 \to H^0(\Gamma, E(\overline{\mathbb{Q}})[n]) \xrightarrow{\iota} H^0(\Gamma, E(\overline{\mathbb{Q}})) \xrightarrow{n} H^0(\Gamma, E(\overline{\mathbb{Q}})) \\ \xrightarrow{\delta} H^1(\Gamma, E(\overline{\mathbb{Q}})[n]) \xrightarrow{\iota} H^1(\Gamma, E(\overline{\mathbb{Q}})) \xrightarrow{n} H^1(\Gamma, E(\overline{\mathbb{Q}})) \end{aligned}$$

where δ is the *connecting homomorphism*. By (3.3), this short exact sequence is the same as

$$0 \to E(\mathbb{Q})[n] \xrightarrow{\iota} E(\mathbb{Q}) \xrightarrow{n} E(\mathbb{Q}) \xrightarrow{\delta} H^1(\Gamma, E(\overline{\mathbb{Q}})[n]) \xrightarrow{\iota} H^1(\Gamma, E(\overline{\mathbb{Q}})) \xrightarrow{n} H^1(\Gamma, E(\overline{\mathbb{Q}})).$$

A proof that this induced sequence holds can be found in most texts on group cohomology and homological algebra e.g. [**4**]. The most difficult part is determining the connecting homomorphism δ, an account of which can be found in [**22**], appendix B.

A corollary of this long exact sequence is below.

Corollary. *There exists a short exact sequence of the groups*

$$0 \to E(\mathbb{Q})/nE(\mathbb{Q}) \to H^1(\Gamma, E(\overline{\mathbb{Q}})[n]) \to H^1(\Gamma, E(\overline{\mathbb{Q}}))[n] \to 0.$$

This short exact sequence is crucial because it characterises $E(\mathbb{Q})/nE(\mathbb{Q})$ from a cohomological viewpoint, something well suited for constructions. It also appears in considerations about existence of rational points on elliptic curves, as well as measuring the failure of Hasse's local-global principle, something discussed in the next section.

3.2 Local Fields

The difficulty of classifying $E(\mathbb{Q})$ is in large part due to the fact $\mathbb{Q}$ is not complete. In other words, the field $\mathbb{Q}$ is not suited to analysis in the same way as $\mathbb{R}$ and $\mathbb{C}$. To overcome this obstacle, we can work with complete field extensions of $\mathbb{Q}$ instead.

Classifying complete field extensions of $\mathbb{Q}$ is not difficult. In fact, in 1916, Ostrowski essentially proved that any completion of $\mathbb{Q}$ is either $\mathbb{R}$ or a p-adic field $\mathbb{Q}_p$. These are the *local fields* we shall work with. A definition of p-adic fields is provided in the appendix.

The idea that the existence of solutions to algebraic curves in local fields implies the existence of rational solutions, is known as *Hasse's Local-Global principle*. It has its roots in the following theorem about quadratic forms.

The Hasse-Minkowski Theorem. *A quadratic form defined over $\mathbb{Q}$ has a non-trivial rational solution if and only if it has a non-trivial solution in $\mathbb{R}$ and every p-adic field $\mathbb{Q}_p$.*

Clearly, one direction of the theorem is obvious since $\mathbb{R}$ and any $\mathbb{Q}_p$ are field extensions of $\mathbb{Q}$. However, the converse is much harder, but also much more useful. A proof can be found in [**6**], chapter 3.

If an algebraic curve contains a solution in every local field, but not in $\mathbb{Q}$, we say that the local-global principle has failed for that curve. Otherwise, we say that the local global principle holds.

By applying Hensel's lemma[1] from p-adic analysis, the Hasse-Minkowski theorem asserts that determining the existence or non-existence of a rational point on a quadratic form amounts to solving congruences.

Of course, one would hope that such a principle would work for algebraic curves that are birationally equivalent to elliptic curves. Unfortunately, this is not always the case. Sometimes it works and sometimes it doesn't. It depends on the elliptic curve in question. As

[1] See Appendix

shall be seen later, the obstruction to the principle depends entirely on the non-triviality of the Tate-Shafarevich group.

This group is inherently cohomological in nature, so it is necessary to reinterpret the existence of rational solutions through a cohomological point of view.

To begin, let E be an elliptic curve defined over $\mathbb{Q}$ and suppose $\Phi : D \longrightarrow E$ is a birational equivalence. Every birational equivalence of E is uniquely determined by an algebraic curve D and a birational map Φ. What's more every such pair (D, Φ) can be associated to a cocycle, namely Λ_Φ. With this in mind, an equivalence relation $\sim$ can be induced on all such birational equivalences of E by the canonical quotient homomorphism

$$q : Z(\Gamma, E(\overline{\mathbb{Q}})) \longrightarrow H^1(\Gamma, E(\overline{\mathbb{Q}})) \ , \ f \mapsto f + B(\Gamma, E(\overline{\mathbb{Q}})).$$

Indeed, let

$$W_E := \ \{(D, \Phi) \mid \Phi : D \longrightarrow E \text{ a birational equivalence}\}$$

be the set of all birational equivalences of E. Then $(D_1, \Phi_1) \sim (D_2, \Phi_2)$ if and only if $\Lambda_{\Phi_1} - \Lambda_{\Phi_2} \in B(\Gamma, E(\overline{\mathbb{Q}}))$. In this case, the two pairs are said to be *cohomologous*.

It follows that the equivalence classes $[D, \Phi] \in W_E/\sim$ map injectively into the first cohomology group $H^1(\Gamma, E(\overline{\mathbb{Q}}))$ via $[D, \Phi] \longrightarrow \Lambda_\Phi + B(\Gamma, E(\overline{\mathbb{Q}}))$. This map is also surjective, see [**22**], chapter 10, meaning that an isomorphic group law can be induced bijectively on $W_E/\sim$. The resulting group is known as the *Weil-Châtelet group* and the identity of this group is denoted [0]. Note that the Weil-Châtelet group can also be defined independently of the first cohomology group via *homogeneous spaces*[2] and the fact that they are isomorphic is entirely non-trivial.

[2] Again, see [**22**], chapter 10 for more.

An important property of the Weil-Châtelet group is summarised by the following lemma.

Lemma 3.2. $[D, \Phi] = [0]$ *if and only if* D *contains a rational point.*

Proof. By (3.3), it is clear that $P \in E(\mathbb{Q})$ if and only if $P \in E(\overline{\mathbb{Q}})$ and $\sigma P = P$ for all $\sigma \in \Gamma$. Now, suppose $[D, \Phi] = [0]$. Since $W_E/\sim$ is naturally isomorphic to $H^1(\Gamma, E(\overline{\mathbb{Q}}))$, this is equivalent to saying the cocycle Λ_Φ is a coboundary (trivial in H^1). Thus,

$$\Lambda_\Phi(\sigma) = \sigma P - P$$

for some $P \in E(\overline{\mathbb{Q}})$.

Re-arranging and using (3.1) this becomes;

$$(\sigma\Phi)\Phi^{-1}(P) = \sigma P.$$

Now, since Φ is a bijection, $P = \Phi(R)$ for some $R \in D$, so the above equation can be rewritten as

$$(\sigma\Phi)(R) = \sigma(\Phi(R)) = (\sigma\Phi)(\sigma R).$$

Hence $\sigma R = R$. Since this holds for all $\sigma \in \Gamma$, R must be rational.

Conversely, let $Q \in D$ be a rational point. Then there exists $P \in E(\overline{\mathbb{Q}})$ such that $P = \Phi(Q)$. Since Q is rational, $\sigma Q = Q$ for all $\sigma \in \Gamma$. Hence,

$$\sigma P = \sigma(\Phi(Q)) = (\sigma\Phi)(Q) = (\sigma\Phi)\Phi^{-1}(P).$$

and thus, (3.1) gives

$$\sigma P = P + \Lambda_\Phi(\sigma).$$

So Λ_Φ is a coboundary. □

This lemma reduces proving rational solutions exist to cohomological considerations. Of course, in this scenario, there is nothing particularly special about $\mathbb{Q}$ and everything discussed so far can be extended to local fields. In particular, using the fact that the first cohomology group and the Weil-Châtelet group are isomorphic, the following lemma holds.

Lemma 3.3. *Let* $\mathbb{F} = \mathbb{R}$ *or some p-adic field* $\mathbb{Q}_p$ *and let* D *be an algebraic curve as before. Then* D *contains a solution in* $\mathbb{F}$ *if and only if there exists a cocycle (corresponding to* D*) which becomes trivial in* $H^1(\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}), E(\overline{\mathbb{F}}))$*. Here* $\overline{\mathbb{F}}$ *denotes the algebraic closure of* $\mathbb{F}$*.*

Lemmas 3.2 and 3.3 are foundational for the Tate-Shafarevich group and its interpretation as an obstacle to a local-global principle.

Next, we will consider how different cohomology groups relate to each other.

By definition, every local field $\mathbb{F}$ is an extension of $\mathbb{Q}$, so the algebraic closure $\overline{\mathbb{F}}$ will be an extension of $\overline{\mathbb{Q}}$. By the process of *localisation*, see [**6**], chapter 21, it is possible to prove that the reverse inclusion holds for the corresponding Galois groups i.e. for every local field $\mathbb{F}$, there is an injective homomorphism

$$\iota_{\mathbb{F}} : \mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}) \longrightarrow \Gamma.$$

Thus, each $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$ can be considered a subgroup of Γ.

There also exists a homomorphism that restricts the domain[3] of every element in $H^1(\Gamma, E(\overline{\mathbb{Q}}))$ to $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$, making it an element of $H^1(\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}), E(\overline{\mathbb{F}}))$. The restriction homomorphism is denoted;

$$\Psi_{\mathbb{F}} : H^1(\Gamma, E(\overline{\mathbb{Q}})) \longrightarrow H^1(\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}), E(\overline{\mathbb{F}})) \tag{3.4}$$

and is defined by taking every element of $H^1(\Gamma, E(\overline{\mathbb{Q}}))$ to its restriction;

$$\Psi_{\mathbb{F}}(f) = f|_{\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})}.$$

With all this in mind, the Tate-Shafarevich group can now be defined.

Let E be an elliptic curve defined over $\mathbb{Q}$ and suppose $\Phi : D \longrightarrow E$ is a birational equivalence, of the algebraic curve D, defined over $\overline{\mathbb{Q}}$. Furthermore, suppose D contains a point in every local field. Then the corresponding cocycle, namely Λ_Φ, is trivial in $H^1(\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p), E(\overline{\mathbb{Q}}_p))$ for every prime p and in $H^1(\mathrm{Gal}(\mathbb{C}/\mathbb{R}), E(\mathbb{C}))$.

[3]Technically, the restriction is to a subgroup of Γ isomorphic to $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$.

Now, the local-global principle would fail for D if we could show that Λ_Φ, considered as an element of $H^1(\Gamma, E(\overline{\mathbb{Q}}))$, was non-trivial in this group. More precisely, if it could be shown that $\Lambda_\Phi \in \mathrm{Ker}(\Psi_\mathbb{F})$ for all local fields $\mathbb{F}$ and is non-trivial in each kernel. The more curves D that share this property, the more the local-global principle fails. Hence, the obstruction to the local-global principle can be measured by the size of the group

$$Ш_E := \bigcap_{\text{Local fields } \mathbb{F}} \mathrm{Ker}(\Psi_\mathbb{F}).$$

This is the Tate-Shafarevich group of E. The dependence on E follows from (3.4).

Clearly, by the above discussion if $Ш_E$ is the trivial group, the local-global principle cannot fail. On the other hand, the larger $Ш_E$ is, the more curves there are that violate the principle. This group is one of the most studied groups in mathematics, but very little is known about it. For most elliptic curves, it is not even known to be finite. It is fundamental to the theory of elliptic curves, not only because it can be used to solve problems about rational points, but also because it plays a vital role in understanding $E(\mathbb{Q})/nE(\mathbb{Q})$.

An equivalent definition of the Tate-Shafarevich group is obtained by taking direct products. Define Ψ to be the product homomorphism obtained from extending each $\Psi_\mathbb{F}$;

$$\Psi : H^1(\Gamma, E(\overline{\mathbb{Q}})) \longrightarrow \prod_{\text{Local fields } \mathbb{F}} H^1(\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}), E(\overline{\mathbb{F}})).$$

Then $Ш_E = \mathrm{Ker}(\Psi)$.

Although $Ш_E$ is not known to be finite, it is still possible to construct finite analogues of this group. These prove essential in bounding the rank of an elliptic curve. For any $n \in \mathbb{N}$, let Ψ_n denote the restriction of Ψ to $H^1(\Gamma, E(\overline{\mathbb{Q}})[n])$

$$\Psi_n : H^1(\Gamma, E(\overline{\mathbb{Q}})[n]) \longrightarrow \prod_{\text{Local fields } \mathbb{F}} H^1(\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}), E(\overline{\mathbb{F}})).$$

Then, an analogue of $Ш_E$ is $S_n := \mathrm{Ker}(\Psi_n)$. This is known as the *nth Selmer group* of E and is *always* finite. The usefulness of these groups comes from the fact they are easily constructable extensions of $E(\mathbb{Q})/nE(\mathbb{Q})$. See [**30**], chapter 8 for an example.

To conclude, consider the short exact sequence from the previous section

$$0 \to E(\mathbb{Q})/nE(\mathbb{Q}) \to H^1(\Gamma, E(\overline{\mathbb{Q}})[n]) \to H^1(\Gamma, E(\overline{\mathbb{Q}}))[n] \to 0.$$

There is nothing special here about $\mathbb{Q}$, so this sequence can be extended to local fields

$$0 \to E(\mathbb{F})/nE(\mathbb{F}) \to H^1(\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}), E(\overline{\mathbb{F}})[n]) \to H^1(\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}), E(\overline{\mathbb{F}}))[n] \to 0.$$

These exact sequences can be rephrased in terms of the Selmer and Tate-Shafarevich groups. A full proof of this can be found in [**22**], chapter 10. However, the basic idea is that the homomorphism $H^1(\Gamma, E(\overline{\mathbb{Q}})[n]) \to H^1(\Gamma, E(\overline{\mathbb{Q}}))[n]$ in the exact sequence above induces a surjective homomorphism between S_n and $Ш_E[n]$. From there, one can apply the snake lemma (a general result in category theory) to the following commutative diagram[4]

$$\begin{array}{ccccccccc}
0 & \longrightarrow & E(\mathbb{Q})/nE(\mathbb{Q}) & \longrightarrow & H^1(\Gamma, E(\overline{\mathbb{Q}})[n]) & \longrightarrow & H^1(\Gamma, E(\overline{\mathbb{Q}}))[n] & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & \overset{\Psi_n}{\searrow} & \downarrow \Psi & & \\
0 & \longrightarrow & \prod_{\mathbb{F}} E(\mathbb{F})/nE(\mathbb{F}) & \longrightarrow & \prod_{\mathbb{F}} H^1(\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}), E(\overline{\mathbb{F}})[n]) & \longrightarrow & \prod_{\mathbb{F}} H^1(\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}), E(\overline{\mathbb{F}}))[n] & \longrightarrow & 0
\end{array}$$

to prove that $E(\mathbb{Q})/nE(\mathbb{Q})$ is isomorphic to the kernel of this homomorphism. In other words, one has the following short exact sequence for each $n \in \mathbb{N}$,

$$0 \to E(\mathbb{Q})/nE(\mathbb{Q}) \to S_n \to Ш_E[n] \to 0.$$

This is one of the most important results in this section and one of the key ideas used in Birch and Swinnerton-Dyer's work. It is commonly known as the *nth descent sequence*.

The existence of an injective homomorphism between $E(\mathbb{Q})/nE(\mathbb{Q})$ and S_n, which will be denoted ψ_n, implies that

$$E(\mathbb{Q})/nE(\mathbb{Q}) \cong \mathrm{Im}(\psi_n) < S_n.$$

With this, the desired isomorphism and generalisation of $E(\mathbb{Q})/2E(\mathbb{Q}) \cong \mathrm{Im}(\psi)$ has been established.

[4]Note that the bottom exact sequence of the commutative diagram follows from taking direct products and product homomorphisms of the short exact local field sequence.

Chapter 4

Birch and Swinnerton-Dyer's 1st Paper

In this chapter, the first of two seminal papers *Notes on elliptic curves I* [**26**] and *Notes on elliptic curves II* [**27**] by Bryan Birch and Peter Swinnerton-Dyer will be discussed. Though the Birch and Swinnerton-Dyer conjecture does not appear until the second paper, examining the first gives an insight into how the conjecture was developed.

The goal of the first paper was the development of an effective algorithm capable of bounding the rank of an elliptic curve E defined over $\mathbb{Q}$. This was accomplished by bounding the corresponding 2-Selmer group S_2 instead, and using the fact that

$$|E(\mathbb{Q})/2E(\mathbb{Q})| \leq |S_2|.$$

Remember that S_2 depends implicitly on the choice of elliptic curve E. It is not a fixed group.

To begin with, Birch and Swinnerton-Dyer reformulated the group S_2 in terms of 2-*coverings*.

Definition 4.1. *Suppose E is an elliptic curve defined over $\mathbb{Q}$ and let* [2] *denote the multiplication-by-*2 *homomorphism. A* 2*-covering of E is a pair $(D, \Phi)_2$, where D is a curve defined over $\mathbb{Q}$ and $\Phi : D \longrightarrow E$ is a birational map defined over $\mathbb{C}$ such that* $[2] \circ \Phi$ *is a rational map defined over $\mathbb{Q}$. We say D admits a* 2*-covering of E.*

Furthemore, a 2*-covering of E, $(D_1, \Phi_1)_2$, is equivalent to another, $(D_2, \Phi_2)_2$, if and only if there is a birational map $\lambda : D_1 \longrightarrow D_2$ defined over $\mathbb{Q}$ and an element of order two in $E(\mathbb{Q})$, denoted T, such that the following diagram commutes*

$$\begin{array}{ccc} D_1 & \xrightarrow{\Phi_1} & E \\ \downarrow{\scriptstyle \lambda} & & \downarrow{\scriptstyle +T} \\ D_2 & \xrightarrow[\Phi_2]{} & E \end{array}$$

The definition of a 2-covering is due to Cassels. See [**5**].

In a similar manner that a group structure can be induced on birational equivalences to form the Weil-Châtelet group, one can induce a group structure on all 2-coverings of E, up to equivalence. Every non-trivial element of this group has order 2. The equivalence class of $(D, \Phi)_2$ will be denoted $[D, \Phi]_2$.

The primary equivalence classes of interest are those $[D, \Phi]_2$ where a representative D contains a solution in every complete field extension of $\mathbb{Q}$ i.e. every local field. Weil [**31**], [**32**] has shown that these equivalence classes form a subgroup isomorphic to S_2. This subgroup is denoted G_E. The fact that this group is isomorphic to S_2 is not entirely surprising. Both groups are finite 2-groups. Both groups depend implicitly on the birational equivalences of elliptic curves. Furthermore, both are defined with reference to local fields.

It follows then from Weil's result that by bounding $|G_E|$, one can bound $|S_2|$ and by extension, $|E(\mathbb{Q})/2E(\mathbb{Q})|$.

Therefore, bounding $|G_E|$ becomes the primary goal. Birch and Swinnerton-Dyer's approach to this was simple.

1. First, they found a canonical form for the curves D, which reduced the number of equivalence class representatives to be considered in G_E.

2. Next, they characterised the equivalence relation of 2-coverings in terms of the canonical form. It followed that every class can be assigned a canonical representative.

3. Finally, they bounded the number of these class representatives. This last step is partially algorithmic in nature.

The first step is resolved through the following lemma;

Lemma 4.1. *If $[D, \Phi]_2 \in G_E$, then it can be assumed without loss of generality that D is of the form*

$$y^2 = g(x) = ax^4 + bx^3 + cx^2 + dx + e \tag{4.1}$$

where $g(x) \in \mathbb{Z}[x]$ and $a \neq 0$.

This lemma is essentially a corollary of the Riemann-Roch theorem. A full proof can be found in the original paper [**26**] while a statement and proof of the Riemann-Roch theorem can be found in [**13**], chapter 4.

Interestingly, a partial converse of this lemma also holds. Every curve of the form (4.1) admits a 2-covering of some elliptic curve. In particular, if one defines

$$I := 12ae - 3bd + c^2,\ J := 72ace + 9bcd - 27ad^2 - 27eb^2 - 2c^3$$

then (4.1) admits a 2-covering (See Definition 4.1) of the elliptic curve

$$y^2 = x^3 - 27Ix - 27J.$$

For example, $D : y^2 = x^4 + 64x^3 + 6x^2 + 1$ admits a 2-covering of

$$y^2 = x^3 - 6^4x + 12^6.$$

The variables I and J are known as the *invariants* of (4.1). This terminology comes from the study of homogeneous polynomials (classically called quantics). Given a polynomial ring $\mathbb{F}[x_1, x_2, \ldots, x_n]$, a polynomial $f \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ is **homogeneous** if every monomial term in f has the same degree. This implies

$$f(\lambda x_1, \lambda x_2, \ldots, \lambda x_n) = \lambda^n f(x_1, x_2, \ldots, x_n) \text{ for all } \lambda \in \mathbb{F}.$$

Suppose $k \in \mathbb{N} \cup \{0\}$. An **invariant** of f is any function ϕ depending only on the coefficients of f, that is invariant (up to a scalar multiple) under a linear transformation of the variables of f. In other words, under the linear map M,

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto M \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \tilde{x_1} \\ \tilde{x_2} \\ \vdots \\ \tilde{x_n} \end{pmatrix}, \; f \mapsto \tilde{f}$$

one has

$$\phi(\tilde{f}) = \det(M)^{2k} \phi(f).$$

One can transform (4.1) into a unique homogeneous polynomial[1] of degree 4 known as its *homogenisation*. The variables I and J are invariants of homogeneous polynomials of degree 4, hence their name. See [**21**], pg 104 and pg 187 - 192 for more on invariants.

A corollary of lemma 4.1 is the fact that every equivalence class in G_E corresponds to a set of quartic curves of the form (4.1). By the definition of 2-coverings, each of these curves must be birationally equivalent over $\mathbb{C}$ to the elliptic curve E.

Our next goal is determining when quartic curves of the form (4.1) admit *equivalent* 2-coverings. To this end, suppose $(D_1, \Phi_1)_2$ and $(D_2, \Phi_2)_2$ are 2-coverings of E where D_1 and D_2 are of the form (4.1).

For equivalence to occur, there must exist a birational map $\lambda : D_1 \longrightarrow D_2$ and this map must satisfy the commutative diagram in definition 3.1 i.e.

$$\Phi_2 \circ \lambda(P) = \Phi_1(P) + T \tag{4.2}$$

for all $P \in D_1$. Here T is an element of order two in $E(\mathbb{Q})$.

[1] See Appendix for an example of how to homogenise a polynomial.

Suppose then D_1 is defined by $y^2 = g_1(x)$ and D_2 is defined by $y^2 = g_2(x)$. An important property that D_1 and D_2 share is that they are both *branched coverings* of the rational projective line $\mathbb{P}_1(\mathbb{Q})$. This essentially means that there exist covering maps from D_1 and D_2 onto a dense subset of the rational projective line[2].

Using this fact, one can induce a bijection $\lambda^* : \mathbb{P}_1(\mathbb{Q}) \longrightarrow \mathbb{P}_1(\mathbb{Q})$ from the birational map $\lambda : D_1 \longrightarrow D_2$. Birch and Swinnerton-Dyer prove that such a map induces the following relation between the polynomials $g_1(x)$ and $g_2(x)$, see [**26**]. There exist $\alpha, \beta, \gamma, \delta, \mu \in \mathbb{Q}$ such that

$$g_1(x) = \mu^2(\gamma x + \delta)^4 g_2\left(\frac{\alpha x + \beta}{\gamma x + \delta}\right). \tag{4.3}$$

This is a re-characterisation of 2-covering equivalence. If two rational quartics g_1, g_2 are related by (4.3), then $y^2 = g_1(x)$, $y^2 = g_2(x)$ admit equivalent 2-coverings of E and vice versa. This completes step two.

Therefore, if one can bound, up to equivalence by (4.3), the number of locally solvable quartic curves $y^2 = g(x)$ that admit 2-coverings of E, this would bound the size of $|G_E|$. To accomplish this, Birch and Swinnerton-Dyer introduced reduced forms for these quartic curves. These reduced forms, R, were defined in such a way that every equivalence class of G_E contained at least one representative of the form $(R, \Phi)_2$. However, reduced forms are not necessarily unique to equivalence classes. It may happen that two distinct reduced forms represent the same class. As Birch and Swinnerton-Dyer mention in their paper, refining them to satisfy uniqueness was not practicable.

For the time being, elliptic curves will be written in the form

$$y^2 = x^3 - 27Ax - 27B. \tag{4.4}$$

This is simply for convenience. Quartic curves that admit 2-coverings of (4.4) have easily deducible invariants, and are therefore easier to work with. Every rational elliptic curve can be written in this way by a suitable scaling of variables. Furthermore, it can be assumed that A and B are integers and that there does not exist a natural number $n > 1$ such that $n^4|A$ and $n^6|B$. Again, these conditions are all a matter of scaling variables correctly.

[2]See Appendix A for more on projective geometry

The first step in defining a reduced form is detailed in the following lemmas. These lemmas provide a partial answer to the question of which quartic curves can admit 2-coverings of (4.4). In the original paper, the first lemma is proved on a case-by-case basis. Here, only a proof of the most general case is provided instead.

Lemma 4.2. *Let $g(x) := ax^4 + bx^3 + cx^2 + dx + e \in \mathbb{Z}[x]$ with invariants I and J. Suppose there exists a prime $p > 4$ such that $p^4|I$, $p^6|J$ and $y^2 = g(x)$ has a solution in $\mathbb{Q}_p$. Then $g(x)$ is equivalent to some quartic $h(x) \in \mathbb{Z}[x]$ with invariants $p^{-4}I$ and $p^{-6}J$.*

Proof. Most general case: Assume $p \nmid \gcd(a, b, c, d, e)$ and $p \nmid b$.

Firstly, note that the solutions to a quartic polynomial $ax^4 + bx^3 + cx^2 + dx + e = 0$ are

$$\begin{aligned} x_{1,2} &= \frac{-b}{4a} - \alpha \pm \frac{1}{2}\sqrt{-4\alpha^2 - 2\rho + \frac{q}{\alpha}}, \\ x_{3,4} &= \frac{-b}{4a} + \alpha \pm \frac{1}{2}\sqrt{-4\alpha^2 - 2\rho - \frac{q}{\alpha}}, \end{aligned}$$

where

$$\rho := \frac{c}{a} - \frac{3}{8}\left(\frac{b}{a}\right)^2, \; q := \frac{d}{a} + \frac{1}{8}\left(\frac{b}{a}\right)^3 - \frac{1}{2}\left(\frac{bc}{a^2}\right)$$

and

$$\alpha^2 = \frac{-1}{6}\rho + \frac{1}{12a}\left(\beta + \frac{I}{\beta}\right), \; (2\beta^3 - J)^2 = J^2 - 4I^3.$$

Now, since $p > 4$ and $\deg(g) = 4$, there must exist some $m \in \mathbb{Z}$ such that $g(m) \not\equiv 0 \bmod(p)$. By using the fact that $g(x)$ and $g(x + m)$ are equivalent (see (4.3)), it can be assumed that $g(0) \not\equiv 0 \bmod(p)$. In other words, $p \nmid e$. Then, using the fact $g(x)$ and $x^4 g(x^{-1})$ are equivalent, this can be re-interpreted as $p \nmid a$. Note that it can no longer be assumed $p \nmid e$.

Next, using the general solution of a quartic polynomial above it can be shown, with some work, that $I \equiv J \equiv 0 \bmod(p)$ implies $g(x) \equiv 0 \bmod(p)$ has a triple root. Using a suitable change in co-ordinates, it will be assumed $g(x) \equiv 0$ has a triple root at zero. What's more, it can also be shown that $p^4|c$. Thus, $g(x) \equiv x^3(ax+b) \bmod(p)$. Hence $p|\gcd(c, d, e)$.

Because $p^2|J$, $p^2|I$ and $p|\gcd(c, d, e)$, it can be deduced that $p^2|e$ and $p^2|d$. Then, from this one can deduce $p^4|d$ and $p^6|e$ since $p^4|I$ and $p^6|J$. Finally, using the birational map $(x, y) \longrightarrow (p^2x, p^3y)$ it follows from (4.3) that the curve

$$y^2 = h(x) := ap^2x^4 + bx^3 + cp^{-2}x^2 + dp^{-4}x + ep^{-6}$$

admits a 2-covering of $y^2 = x^3 - 27Ix - 27J$ that is equivalent to the 2-covering $y^2 = g(x)$. The invariants of $y^2 = h(x)$ are $p^{-4}I$ and $p^{-6}J$. Thus, $g(x)$ is equivalent to some $h(x) \in \mathbb{Z}[x]$, which has invariants $p^{-4}I$ and $p^{-6}J$. □

In the circumstances where $p = 2$ or $p = 3$, the following lemmas hold. The expression $p^k \parallel n$ means $p^k \mid n$ but $p^{k+1} \nmid n$.

Lemma 4.3. *Let $g(x) := ax^4 + bx^3 + cx^2 + dx + e \in \mathbb{Z}[x]$ with invariants I and J. Suppose $2^6|I$, $2^9|J$ and $2^{10}|(8I + J)$. If $y^2 = g(x)$ has a solution in $\mathbb{Q}_2$ then $g(x)$ is equivalent to some quartic $h(x) \in \mathbb{Z}[x]$ with invariants $2^{-4}I$ and $2^{-6}J$.*

Lemma 4.4. *Let $g(x) := ax^4 + bx^3 + cx^2 + dx + e \in \mathbb{Z}[x]$ with invariants I and J. Suppose $3^5|I$ and $3^9|J$ or $3^4 \parallel I$, $3^6 \parallel J$ and $3^{15}|(4I^3 - J^2)$. If $y^2 = g(x)$ has a solution in $\mathbb{Q}_3$ then $g(x)$ is equivalent to some quartic $h(x) \in \mathbb{Z}[x]$ with invariants $3^{-4}I$ and $3^{-6}J$.*

The proofs of these lemmas are similar to lemma 4.2, but are much longer.

Given invariants I and J of a 2-covering $y^2 = g(x)$, an immediate consequence of these three lemmas is that in a significant number of cases, it can be assumed that there does not exist an integer $n > 1$ such that $n^4|I$ and $n^6|J$. This reduces the work required in classifying the 2-coverings of (4.4).

These lemmas conclude the first step toward defining reduced quartic curves. The second (main) step involves concepts from the classical theory of invariants and late 19th century algebra, see [**21**] for more.

To begin, several definitions from the theory of quadratic forms are provided.

Definition 4.2. *A binary quadratic form is a polynomial of the form;*

$$f(x, y) = ax^2 + bxy + cy^2$$

where $a, b, c \in \mathbb{Z}$. Equivalently, a binary quadratic form is a homogeneous integral polynomial of degree two.

As with quartic polynomials, there is a notion of equivalence amongst binary quadratic forms. Notice the similarity to (4.3).

Definition 4.3. *The quadratic forms $f(x, y) = ax^2 + bxy + cy^2$ and $g(x, y) = \tilde{a}x^2 + \tilde{b}xy + \tilde{c}y^2$ are **equivalent** if and only if there exist integers $\alpha, \beta, \gamma, \delta$ such that*

$$f(x, y) = (\gamma x + \delta y)^2 g\left(\frac{\alpha x + \beta y}{\gamma x + \delta y}, 1\right) = g(\alpha x + \beta y, \gamma x + \delta y), \text{ where } \alpha\delta - \beta\gamma = 1. \quad (4.5)$$

Particular attention is given to two special types of binary quadratic forms, which are defined below.

Definition 4.4. *A quadratic form $f(x, y) = ax^2 + bxy + cy^2$ is said to be positive definite if $a > 0$ and $b^2 - 4ac < 0$. These conditions are necessary and sufficient for $f(x, y)$ to obtain only positive values over the real numbers.*

Definition 4.5. *A form $f(x, y) = ax^2 + bxy + cy^2$ is said to be reduced if it is positive definite, $|b| \leq a \leq c$ and $b \geq 0$ if $a = |b|$ or $a = c$.*

An interesting theorem that explains how the previous definitions relate to each other is the following.

Theorem 4.1. *Suppose $f(x, y) = ax^2 + bxy + cy^2$ is positive definite and $(a, b, c) = 1$, then $f(x, y)$ is equivalent to a unique reduced form.*

Birch and Swinnerton-Dyer wanted to establish a similar but weaker result for quartic curves and 2-coverings.

To accomplish this, they utilised ideas from Julia's PhD thesis [**16**]. Part of Julia's thesis focused on the equivalence of binary quartic forms (degree 4 homogeneous polynomials) under real linear transforms (similar to (4.5)). To each quartic form, Julia associated a positive definite quadratic form. He defined the quartic form to be *reduced* if the associated quadratic form was, in which case the quartic form would act as a unique class representative under the equivalence mentioned above.

The positive definite quadratic form associated to the quartic form was a *covariant*, defined below.

Definition 4.6. *Given a polynomial,*

$$f(x_1, x_2, \ldots, x_n) = \sum_{i_1,\ldots,i_n} \alpha_{i_1,\ldots,i_n} x_1^{i_1}, \ldots, x_n^{i_n}$$

*a **covariant** of f is a polynomial $\Phi(\alpha, x_1, x_2, \ldots, x_n)$ depending on the variables and coefficients of f (collectively called α) such that under any linear transform M*

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto M \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} x_1^{'} \\ x_2^{'} \\ \vdots \\ x_n^{'} \end{pmatrix}$$

one has

$$\Phi(\alpha', x_1^{'}, x_2^{'}, \ldots, x_n^{'}) = \det(M)^k \Phi(\alpha, x_1, x_2, \ldots, x_n)$$

for some $k \in \mathbb{N} \cup \{0\}$. The variable α' denotes the coefficients of $f(x_1^{'}, x_2^{'}, \ldots, x_n^{'})$.

Consider the quartic form

$$f(x, y) = ax^4 + 4bx^3y + 6cx^2y^2 + 4dxy^3 + ey^4.$$

An example of a covariant of $f(x, y)$ is

$$\Phi = (ax^2 + 2bxy + cy^2)(cx^2 + 2dxy + ey^2) - (bx^2 + 2cxy + dy^2)^2.$$

The function Φ is also a multiple of the Hessian determinant of $f(x, y)$. This description might make it easier to see why Φ is a covariant. For more on covariants see [**21**], pg 107, pg 126-129 and pg 169.

Birch and Swinnerton-Dyer defined reduced quartic curves $y^2 = g(x)$ in a weaker way than Julia defined reduced quartic forms. The simplest way of defining a reduced quartic curve would have been to follow Julia and define $y^2 = g(x)$ to be reduced if one could find a quadratic covariant of $g(x)$, say[3] $h(x) = \alpha x^2 + \beta x + \gamma$, whose homogeneous form/homogenisation $\tilde{h}(x, y) = \alpha x^2 + \beta xy + \gamma y^2$ is a reduced quadratic form. This, in particular, would have implied

[3]The dependence of h on the coefficients of $g(x)$ is implicit.

1. There exists a covariant of $g(x)$, $\alpha x^2 + \beta x + \gamma$, such that $\alpha^2 \leq \frac{1}{3}\left(\beta^2 - 4\alpha\gamma\right)$.

2. If $g(x) = ax^4 + bx^3 + cx^2 + dx + e$, then one can assume $-2|a| \leq b \leq 2|a|$.

However, this is not the optimum way of defining a reduced quartic curve. A better approach involves using inequalities that are weaker than those mentioned above.

Consider the following: rewrite $g(x)$ as

$$g(x) = a(x^2 + px + q)(x^2 + p'x + q').$$

This can always be done by applying the conjugate roots theorem. It follows that

$$\begin{aligned} a(p + p') &= b, & 3a(q + q') &= c + \phi, \\ 3a(pp') &= 2c - \phi, & a(qq') &= e, \\ a(pq' + p'q) &= d, \end{aligned}$$

where ϕ is a real root of the cubic $x^3 - 3Ix + J$. This cubic has discriminant $4I^3 - J^2$. Here I and J are the invariants of $g(x)$. In the case where $x^3 - 3Ix + J$ has three real distinct roots (in other words, $4I^3 - J^2 > 0$), they shall be labelled $\phi_1 > \phi_2 > \phi_3$ or $\phi_1 < \phi_2 < \phi_3$ according to whether $a > 0$ or $a < 0$. If $g(x)$ has no real roots, it is assumed $a > 0$, otherwise $y^2 = g(x)$ would have no solutions. If $x^3 - 3Ix + J$ has only one real root (implying $4I^3 - J^2 < 0$), it shall simply be denoted ϕ.

With the above notation in mind, Birch and Swinnerton-Dyer construct covariants for assumption 1 and prove that the inequalities in 1 and 2 above imply the following theorems.

Theorem 4.2. *Suppose $g(x) = ax^4 + bx^3 + cx^2 + dx + e \in \mathbb{Z}[x]$ has no real roots and define $K := \frac{1}{3}(4I - \phi_1^2)$. Then under the inequalities in assumptions* 1 *and* 2 *above, one has*

1. $-2|a| \leq b \leq 2|a|$,

2. $8ac - 3b^2 \leq 2K + 2\sqrt{K}(\phi_1 - 3a) - 2a\phi_1$.

Theorem 4.3. *Suppose* $g(x) = ax^4 + bx^3 + cx^2 + dx + e \in \mathbb{Z}[x]$ *has two real roots. Then under the inequalities in assumptions* 1 *and* 2 *above, one has*

1. $-2|a| \leq b \leq 2|a|$,
2. $8ac - 3b^2 \geq 9a^2 - 2a\phi + \frac{1}{3}(4I - \phi^2)$.

Theorem 4.4. *Suppose* $g(x) = ax^4 + bx^3 + cx^2 + dx + e \in \mathbb{Z}[x]$ *has four real roots. Then under the inequalities in assumptions* 1 *and* 2 *above, one has*

1. $-2|a| \leq b \leq 2|a|$,
2. $8ac - 3b^2 \geq 4a\phi_2 - \frac{4}{3}(I - \phi_2^2)$.

The inequalities in theorems 4.2, 4.3 and 4.4 are not only weaker than the inequalities in assumptions 1 and 2 above, but are still sufficiently strong for an analogous form of theorem 4.1 to hold for quartic curves. As a result, these inequalities along with lemmas 4.2, 4.3 and 4.4 are the results that Birch and Swinnerton-Dyer based the following definition of reduced quartic curves on.

Definition 4.7. *The quartic curve* $y^2 = g(x) := ax^4 + bx^3 + cx^2 + dx + e \in \mathbb{Z}[x]$ *with* $a \neq 0$ *and invariants I, J is said to be* ***reduced*** *if both* 1 *and* 2 *below are satisfied and one of* 3, 4 *or* 5 *is also satisfied:*

1. *There is no prime* $p > 4$ *such that* $p^4|I$ *and* $p^6|J$.
2. *It is not true that* $3^5|I$ *and* $3^9|J$, *nor is it true that* $3^4 \parallel I$, $3^6 \parallel J$ *and* $3^{15}|(4I^3 - J^2)$, *nor that* $2^6|I$, $2^9|J$ *and* $2^{10}|(8I + J)$.
3. *the quartic* $g(x)$ *has no real roots,* $a > 0$ *and the inequalities in theorem* 4.2 *hold.*
4. *the quartic* $g(x)$ *has two real roots and the inequalities in theorem* 4.3 *hold.*
5. *the quartic* $g(x)$ *has four real roots and the inequalities in theorem* 4.4 *hold.*

As already stated, this definition was chosen because it was sufficient for the following theorem to hold.

Theorem 4.5. *Let E be a rational elliptic curve and* $[D, \Phi]_2 \in G_E$. *Then there exists a reduced quartic curve* $R : y^2 = g(x)$ *such that R admits an equivalent*[4] 2*-covering as D.*

Thus, every 2-covering in G_E can be represented by a reduced quartic curve. This theorem concludes the first half of the paper.

[4] See definition 4.1

The second half of the paper is dedicated to proving the following

Theorem 4.6. *The reduced quartic curves that admit* 2*-coverings of E are finite in number and computable.*

According to Birch and Swinnerton-Dyer, the finiteness of these reduced quartic curves is implicit in the following algorithm they used to compute them.

Due to the relatively slow processing power available during the 1960s, developing an *effective* algorithm for computing reduced quartic curves was difficult. Even with the EDSAC-2, the advanced computer system that Birch and Swinnerton-Dyer used, they were only able to compute several thousand cases.

The algorithm they developed for computing these reduced quartic curves can be summarised by the following steps:

Step 1: Using the definition of reduced quartic curves, determine the permissible values of the invariants I and J and then list all reduced curves with said invariants that admit 2-coverings of E.

This part of the algorithm takes up most of the computer time needed for the entire process, in large part due to the fact this step makes extensive use of the inequalities in theorems 4.2, 4.3 and 4.4. Birch and Swinnerton-Dyer describe how some leeway in solving these inequalities must be accounted for due to the fact that trial values are generally not integers. Small round-off errors in the programming may cause these inequalities to appear incorrect. In practise, this leeway leads to extra reduced curves being listed.

Step 2: Reduced curves that are equivalent to a curve already obtained are rejected. Furthermore, curves that are birational to E over $\mathbb{Q}$ are rejected. These admit trivial 2-coverings.

The first part of this step amounts to using (4.3) to determine whether two reduced curves are equivalent. The second part of this step rejects a reduced curve $y^2 = g(x)$ if $g(x)$ has a rational root. Curves of this form are birationally equivalent to E over $\mathbb{Q}$, see [**30**], pg 37.

Step 3: Reject the remaining reduced curves that have no solutions in some p-adic field $\mathbb{Q}_p$.

This step requires the use of two lemmas

Lemma 4.5. *Suppose $y^2 = g(x)$ is a reduced quartic curve with invariants I and J where $\nu_p(4I^3 - J^2) > 0$. For a given prime $p > 2$ and a given $x_0 \in \mathbb{Q}_p$, define $\lambda := \nu_p(g(x_0))$ and $\mu := \nu_p(g'(x_0))$. Then $y^2 = g(x)$ has a p-adic solution (x_1, y_1), where $p^\eta | (x_1 - x_0)$, if $g(x_0) = \alpha^2$ for some $\alpha \in \mathbb{Q}_p$ or if $\lambda - \mu \geq \eta > \mu$.*

and

Lemma 4.6. *Suppose $y^2 = g(x)$ is a reduced quartic curve with invariants I and J where $\nu_2(4I^3 - J^2) > 0$. For a given $x_0 \in \mathbb{Q}_2$, define $\lambda := \nu_2(g(x_0))$ and $\mu := \nu_2(g'(x_0))$. Then $y^2 = g(x)$ has a 2-adic solution (x_1, y_1), where $2^\eta | (x_1 - x_0)$, if $g(x_0)$ is a 2-adic square, or if $\lambda - \mu \geq \eta > \mu$, or if $\eta > \mu$ and $\lambda = \mu + \eta - 1$ is even, or if $\eta > \mu$, $\lambda = \mu + \eta - 2$ is even and $2^{-\lambda} g(x_0) \equiv 1 \mod 4$.*

These lemmas, along with Hensel's lemma provide a method by which a computer can determine if p-adic solutions exist. This is accomplished by exhausting a list of possible congruences. If a solution to one of these congruences exist, a p-adic solution exists and the computation is finished. On the other hand, if the computer exhausts all possible congruences and no solutions are found, we assume the curve has no solution in $\mathbb{Q}_p$.

Finally, by computing the number of reduced curves that remain, one immediately obtains a bound on the rank of E. From there, it is then possible to check whether this bound is exact.

The next several pages include some of the results Birch and Swinnerton-Dyer obtained in their paper.

In the tables below, g denotes the rank of an elliptic curve E, while t is a power denoting the cardinality

$$\left|Ш_E[2]\right| := 2^t.$$

Given elliptic curves of the form $E_D : y^2 = x^3 - Dx$ and $E'_D : y^2 = x^3 - D$, the following tables list values of D for which the rank g and the power t are fixed.

Table 2

Curves $y^2 = x^3 - D$ for which g and t take given values.

a) Curves with $g = t = 0$.

1	3	5	6	8	9	10	12	14	16	17	24	27
31	32	33	34	36	37	41	42	46	52	62	68	69
70	73	77	78	80	82	86	88	90	92	96	97	98
99	103	108	111	114	117	119	125	132	133	134	136	144
145	154	156	158	160	161	162	168	169	176	177	178	181
183	189	190	194	196	197	205	206	208	213	217	220	221
224	226	227	230	232	240	241	247	248	250	255	257	258
260	263	266	269	275	276	278	280	283	285	296	302	304
305	306	312	313	314	315	319	321	330	333	335	338	340
341	349	351	352	358	360	361	374	376	377	378	380	385
386	388	392	394	396	399	400	1005	1008	1009			
—1	—4	—6	—7	—13	—14	—16	—20	—21	—23	—25	—27	—29
—32	—34	—42	—45	—49	—51	—53	—59	—60	—70	—75	—78	—81
—84	—85	—86	—87	—88	—90	—93	—95	—96	—104	—109	—114	—115
—116	—124	—125	—135	—137	—140	—144	—153	—157	—158	—159	—160	—162
—165	—167	—173	—175	—176	—178	—180	—181	—187	—193	—194	—200	—201
—203	—209	—216	—228	—230	—237	—239	—240	—242	—243	—244	—245	—250
—253	—258	—259	—261	—266	—267	—270	—273	—279	—284	—292	—300	—301
—302	—304	—305	—306	—308	—309	—311	—312	—317	—324	—330	—338	—339
—341	—345	—348	—358	—361	—363	—367	—372	—374	—375	—376	—378	—387
—394	—397	—400	—1008									

b) Curves with $g = 1$, $t = 0$.

2	4	7	13	15	18	19	20	21	22	23	25	28
29	30	35	38	40	43	44	45	48	49	50	51	54
55	56	57	58	59	60	63	65	66	71	72	74	75
79	81	84	85	87	91	93	94	95	100	101	102	107
110	112	115	120	123	124	126	127	129	130	131	135	137
138	140	143	146	148	150	151	153	157	159	163	164	165
166	167	171	172	173	175	179	180	182	184	187	188	193
195	199	201	202	203	204	209	210	215	216	218	223	225
228	229	231	234	235	237	238	239	242	243	245	246	251
252	254	259	261	264	265	267	268	270	271	272	273	274
279	281	282	284	287	288	290	292	295	297	300	301	303
308	309*	310	311	316	317	318	323	324	325	326	328	331
332	336	337	339	342	343	344	345	347	348	353	354	359
363	367	369	372	373	375	379	381	382	383	387	389	390
395	398	1001	1007	1010								
—2	—3	—5	—8	—9	—10	—11	—12	—18	—19	—22	—26	—28
—30	—31	—33	—35	—36	—38	—39	—40	—41	—44	—46	—47	—48
—50	—52	—54	—55	—56	—58	—61	—62	—66	—67	—68	—69	—71
—72	—74	—76	—77	—80	—82	—83	—91	—92	—94	—97	—98	—99
—100	—102	—103	—105	—107	—108	—110	—111	—112	—117	—118	—119	—120
—121	—126	—127	—130	—132	—133	—134	—136	—138	—139	—143	—146	—147
—149	—150	—152	—154	—155	—156	—163	—166	—169	—170	—172	—177	—179
—182	—183	—184	—185	—188	—189	—190	—191	—196	—199	—202	—205	—206
—207	—208	—210	—211	—212	—213	—215	—218	—219	—220	—221	—224	—226
—227	—234	—235	—236	—238	—241	—246	—247	—249	—251	—254	—255	—257
—262	—263	—264	—271	—272	—274	—275	—276	—277	—278	—280	—282	—283
—285	—287	—288	—289	—290	—291	—293	—296	—298	—299	—307	—310	—313
—314	—315	—318	—319	—321	—323	—325	—326	—327	—328	—329	—332	—333
—334	—335	—336	—340	—342	—343	—344	—349	—351	—352	—354	—355	—356
—357	—364	—365	—368	—370	—371	—379	—380	—382	—385	—386	—390	—391
—393	—395	—396	—398	—399	—1002	—1003	—1005	—1007	—1010			

c) Curves with $g = 2$, $t = 0$.

11	26	39	47	53	61	67	76	83	89	104	106	109
116	118	121	139	147	152	155	170	186	191	200	207	211
212	214	219	222	233	236	244	249	262	277	286	289	291
293	294	298	299	327	329	334	350	355	356	364	366	368
370	371	391	393	397	1003							
—15	—17	—24	—37	—43	—57	—63	—65	—73	—79	—89	—101	—106
—122	—129	—131	—142	—145	—148	—151	—161	—164	—168	—171	—186	—195
—197	—198	—204	—217	—222	—223	—225	—229	—232	—233	—248	—252	—260
—265	—268	—269	—281	—294	—295	—297	—303	—322	—331	—337	—347	—350
—353	—360	—366	—369	—373	—377	—381	—388	—389	—392	—1001	—1004	—1006
—1009												

d) Curves with $g = 3$, $t = 0$.

174	307	362		
—113	—141	—316	—346	—359

e) Curves with $g = 0$, $t = 2$.

105	113	122	141	142	149	185	198	253	322	357	365	1006
—123	—214	—231	—286	—383								

f) Curves with $g = 1$, $t = 2$.

346	
—174	—362

Table 3

Curves $y^2 = x^3 - Dx$ with positive D for which g and t take given values.

a) Curves with $g = t = 0$.

1	3	4	8	9	11	13	18	19	24	27	28	29
33	35	40	43	44	51	59	61	63	67	68	75	83
88	91	92	93	98	100	104	107	108	109	115	120	121
123	125	126	129	131	139	152	153	157	163	164	168	172
173	177	179	180	187	189	195	198	200	228	388	468	548
644	708	772										

b) Curves with $g = 1$, $t = 0$.

2	5	6	7	10	12	14	15	20	21	22	23	25
26	30	31	34	36	37	38	39	41	42	45	46	47
49	50	52	53	54	55	57	58	60	62	66	69	70
71	72	73	74	76	78	79	84	85	86	87	89	94
95	99	101	102	103	105	106	110	111	114	116	118	119
122	124	127	130	133	134	135	137	138	140	142	143	146
147	148	149	150	151	154	158	159	165	166	167	169	170
171	174	175	178	181	182	183	185	186	188	190	191	194
196	197	199										

c) Curves with $g = 2$, $t = 0$.

17	56	65	77	90	97	117	132	136	141	145	155	156
161	184											

d) Curves with $g = 3$, $t = 0$.

82

e) Curves with $g = 0$, $t = 2$.

113	193	248	328	568	584	776

Table 4

Curves $y^2 = x^3 - Dx$ with negative D for which g and t take given values.

a) Curves with $g = t = 0$.

—1	—2	—4	—6	—7	—10	—11	—12	—22	—23	—25	—26	—27
—30	—36	—38	—42	—43	—44	—45	—50	—52	—54	—58	—59	—70
—71	—72	—74	—75	—76	—78	—86	—87	—91	—102	—103	—106	—107
—108	—110	—116	—118	—119	—122	—123	—130	—132	—134	—135	—139	—140
—147	—151	—166	—167	—169	—170	—172	—174	—182	—186	—187	—190	—199

b) Curves with $g = 1$, $t = 0$.

—3	—5	—8	—9	—13	—15	—18	—19	—20	—21	—24	—28	—29
—31	—35	—37	—40	—47	—49	—51	—53	—56	—60	—61	—67	—69
—77	—79	—83	—84	—85	—88	—90	—92	—93	—95	—98	—100	—101
—104	—109	—111	—115	—120	—121	—124	—125	—126	—127	—131	—133	—141
—143	—148	—149	—152	—153	—156	—157	—159	—163	—164	—165	—168	—171
—173	—175	—179	—180	—181	—184	—188	—189	—191	—195	—196	—197	—198
—200												

c) Curves with $g = 2$, $t = 0$.

—14	—33	—34	—39	—46	—55	—63	—65	—66	—68	—73	—89	—94
—99	—105	—113	—114	—129	—138	—145	—150	—154	—155	—158	—178	—183
—185												

e) Curves with $g = 0$, $t = 2$.

—17	—41	—57	—62	—82	—97	—117	—137	—142	—146	—161	—177	—193
—194	—392	—452	—792									

f) Curves with $g = 1$, $t = 2$.

—136	—584	—712	—776

Chapter 5

Preliminaries

Before continuing the discussion of Birch and Swinnerton-Dyer's work, it is necessary to introduce some preliminary notions. This chapter includes several definitions and results that will be needed later on to understand their conjecture. A complete account of everything discussed in this chapter can also be found in [**30**].

5.1 Finite Fields

At its core, the Birch and Swinnerton-Dyer conjecture is a statement about the ranks of rational elliptic curves. Heuristically, the conjecture asserts that the cardinalities of finite field elliptic curves, $E(\mathbb{F}_q)$, determine the ranks of rational elliptic curves $E(\mathbb{Q})$. In section 2.4 it was specified that an arbitrary finite field elliptic curve is not of much interest on its own. However, what is interesting is how they behave in general.

Firstly, some definitions are necessary.

Definition 5.1. *Let $\overline{\mathbb{F}}$ be the algebraic closure of a field $\mathbb{F}$ and E be an elliptic curve defined over this closure. An **endomorphism** of E is a homomorphism $\alpha : E(\overline{\mathbb{F}}) \longrightarrow E(\overline{\mathbb{F}})$ given by rational functions in $\overline{\mathbb{F}}$;*

$$\alpha(x, y) = (R_1(x, y), R_2(x, y))$$

where $R_1(x, y), R_2(x, y) \in \overline{\mathbb{F}}(x, y)$.

The trivial homomorphism, which sends every element to the identity is also considered to be an endomorphism. It is denoted 0. Like homomorphisms, endomorphisms form a ring under addition and composition.

It is not difficult to show that every endomorphism of an elliptic curve is of the form

$$\alpha(x, y) = \left(\frac{p_1(x)}{q_1(x)}, \frac{p_2(x)}{q_2(x)}y\right)$$

where $p_i(x), q_i(x) \in \overline{\mathbb{F}}[x]$ for $i \in \{1, 2\}$ and it is assumed that $\alpha(x_0, y_0) = \infty$ if $q_1(x_0) = 0$ (which implies $q_2(x_0) = 0$). This formula follows from cases 3 and 4 of the Addition Law.

The **degree** of an endomorphism α describes how complicated its rational components are. It is defined as

$$\deg(\alpha) := \max\{\deg(p_1), \deg(q_1)\}$$

When $\alpha = 0$, the assumption is that $\deg(\alpha) = 0$.

An endomorphism α is defined to be **separable** if the formal derivative of $\frac{p_1(x)}{q_1(x)}$ is not identically zero. Analogous to separability in Galois theory, it indicates distinctness of polynomial roots.

The usefulness of these definitions is exemplified by the following theorem.

Theorem 5.1. *If $\alpha \neq 0$ is a separable endomorphism of E, then*

$$\deg(\alpha) = |\mathrm{Ker}(\alpha)|$$

and if $\alpha \neq 0$ is not separable, then

$$\deg(\alpha) > |\mathrm{Ker}(\alpha)|$$

A proof of this theorem can be found in [**30**], chapter 2.

The particular endomorphism that we are interested in is known as the *Frobenius endomorphism*.

Definition 5.2. *Given a finite field $\mathbb{F}_q$ and an elliptic curve E which is defined over $\mathbb{F}_q$, the **Frobenius endomorphism** is defined as the map*

$$\phi_q : E(\overline{\mathbb{F}_q}) \longrightarrow E(\overline{\mathbb{F}_q}),\ \phi_q(x, y) \mapsto (x^q, y^q)$$

By assumption, $\phi_q(\infty) := \infty$.

Proving this map is an endomorphism is not difficult. It is simply a matter of using the Addition Law formulas and the fact $x \longrightarrow x^q$ is a ring homomorphism of $\overline{\mathbb{F}_q}$. It is immediate from the definition of the Frobenius endomorphism that it is not separable and has degree q.

The Frobenius endomorphism can be used to determine the elements of $E(\mathbb{F}_q)$ in the following way.

Lemma 5.1. *Suppose* $(x, y) \in E(\overline{\mathbb{F}_q})$*, then* $(x, y) \in E(\mathbb{F}_q)$ *if and only if* $\phi_q(x, y) = (x, y)$*. Therefore,* $\mathrm{Ker}(\phi_q - 1) = E(\mathbb{F}_q)$*, where* 1 *denotes the identity homomorphism.*

This lemma is an immediate corollary of the fact $a \in \mathbb{F}_q$ if and only if $a \in \overline{\mathbb{F}_q}$ and $a^q - a = 0$.

If one could prove that the endomorphism $\phi_q - 1$ was separable, it would be possible to relate the degree of $\phi_q - 1$ to the cardinality of $E(\mathbb{F}_q)$ by the use of theorem 5.1. To this end, the following result is required.

Theorem 5.2. *Suppose* $\mathbb{F}_q$ *is a finite field of characteristic* p *and* E *is an elliptic curve defined over* $\mathbb{F}_q$*. Furthermore, let* $a, b \in \mathbb{Z}$*, not both zero, denote the multiplication-by-a and multiplication-by-b homomorphisms respectively. Then* $a\phi_q + b$ *is a separable endomorphism if and only if* $p \nmid b$*.*

A proof can be found in [**30**], chapter 2.

By choosing $a = 1$ and $b = -1$, it is obvious from this theorem that $\phi_q - 1$ is separable. Thus,

$$|E(\mathbb{F}_q)| = |\mathrm{Ker}(\phi_q - 1)| = \deg(\phi_q - 1). \tag{5.1}$$

In order to determine (or at least estimate) the cardinality of $E(\mathbb{F}_q)$, the following theorem about degrees is also necessary. A partial proof can be found in [**30**], chapter 3.

Theorem 5.3. *Let* α, β *be two endomorphisms of* $E(\overline{\mathbb{F}_q})$ *and* $a, b \in \mathbb{Z}$ *denote the multiplication-by-a and multiplication-by-b homomorphisms, then*

$$\deg(a\alpha + b\beta) = a^2 \deg(\alpha) + b^2 \deg(\beta) + ab(\deg(\alpha + \beta) - \deg(\alpha) - \deg(\beta)). \tag{5.2}$$

By choosing $\alpha = \phi_q$ and $\beta = -1$, (5.2) becomes

$$\deg(a\phi_q - b) = a^2 q + b^2 - ab(q + 1 - \deg(\phi_q - 1)).$$

Now, $\deg(a\phi_q - b) \geq 0$ by definition, thus,

$$a^2 q + b^2 - ab(q + 1 - \deg(\phi_q - 1)) \geq 0. \tag{5.3}$$

For any non-zero a, dividing across (5.3) by a^2 gives

$$\left(\frac{b}{a}\right)^2 - (q + 1 - \deg(\phi_q - 1))\left(\frac{b}{a}\right) + q \geq 0$$

Since $\mathbb{Q}$ is dense in $\mathbb{R}$ and $\frac{b}{a}$ can be an arbitrary rational, the polynomial $x^2 - (q + 1 - \deg(\phi_q - 1))x + q$ must be non-negative for all $x \in \mathbb{R}$. Thus, it must have a non-positive discriminant

$$(q + 1 - \deg(\phi_q - 1))^2 - 4q \leq 0.$$

Applying (5.1), this is equivalent to

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

This is known as **Hasse's theorem**. See [**14**]. A corollary of this result that is relevant to the Birch and Swinnerton-Dyer conjecture is the following.

Corollary. $\lim_{q\to\infty}\left(\frac{|E(\mathbb{F}_q)|}{q}\right) = 1.$

The next result, which will also be relevant later, tells us how to relate the cardinality of $E(\mathbb{F}_q)$ to its extension $E(\mathbb{F}_{q^n})$ for any $n \in \mathbb{N}$. Firstly, it can be shown, see [**30**], that using the Cayley-Hamilton theorem and the fact that endomorphisms form a ring under addition and composition, that the following equation is true for $a = q + 1 - \deg(\phi_q - 1)$, namely

$$\phi_q^2 - a\phi_q + q = 0. \tag{5.4}$$

Furthermore, $a = q + 1 - \deg(\phi_q - 1)$ is the unique integer for which this equation holds. Note that if $x^2 - ax + q = (x - \alpha)(x - \beta) \in \mathbb{Z}[x]$, where $\alpha, \beta \in \mathbb{C}$, then clearly $a = \alpha + \beta$ and by (5.1), $|E(\mathbb{F}_q)| = q + 1 - \alpha - \beta$. This result can be generalised as follows.

Theorem 5.4. *Let* $a = q + 1 - \deg(\phi_q - 1)$ *and* $x^2 - ax + q = (x - \alpha)(x - \beta)$*, then* $|E(\mathbb{F}_{q^n})| = q^n + 1 - \alpha^n - \beta^n$.

Proof. Clearly, $\alpha, \beta \in \mathbb{C}$, so the first priority is showing $\alpha^n + \beta^n$ is indeed an integer.

Clearly, $\alpha^0 + \beta^0 = 2$ and $\alpha^1 + \beta^1 = a$ are integers. Since α and β are roots of $x^2 - ax + q$, it's easily seen that

$$\begin{aligned} \alpha^{n+1} &= a\alpha^n - q\alpha^{n-1} \\ \beta^{n+1} &= a\beta^n - q\beta^{n-1}. \end{aligned}$$

Therefore,

$$\alpha^{n+1} + \beta^{n+1} = a(\alpha^n + \beta^n) - q(\alpha^{n-1} + \beta^{n-1}).$$

Hence, by induction, $\alpha^n + \beta^n$ is an integer for any $n \in \mathbb{N}$.

Now consider the polynomial $(x^n - \alpha^n)(x^n - \beta^n) = x^{2n} - (\alpha^n + \beta^n)x^n + q^n \in \mathbb{Z}[x]$. Clearly, $(x-\alpha)(x-\beta) = x^2 - ax + q$ divides this polynomial. Therefore, using (5.4) and evaluating at ϕ_q one has

$$\phi_q^{2n} - (\alpha^n + \beta^n)\phi_q^n + q^n = 0.$$

However, ϕ_q composed with itself n times is just ϕ_{q^n}. In other words, $\phi_q^n = \phi_{q^n}$. Thus,

$$\phi_{q^n}^2 - (\alpha^n + \beta^n)\phi_{q^n} + q^n = 0.$$

If we denote $q' = q^n$, then (5.4) says that $\phi_{q'}$ must satisfy

$$\phi_{q'}^2 - a'\phi_{q'} + q' = 0$$

where $a' = q' + 1 - \deg(\phi_{q'} - 1)$ is the unique integer for which this equation holds.

Therefore, $\alpha^n + \beta^n = a' = q' + 1 - \deg(\phi_{q'} - 1) = q^n + 1 - |E(\mathbb{F}_{q^n})|$.

□

In the following short section, aspects from the theory of reduction are discussed with a view toward linking rational elliptic curves and finite field elliptic curves.

5.2 Reduction

Suppose $E : y^2 = x^3 + Ax + B$ is an elliptic curve defined over $\mathbb{Q}$. Without loss of generality, it can be assumed that $A, B \in \mathbb{Z}$ by scaling variables correctly. For an arbitrary prime p, this elliptic curve can be mapped into $\mathbb{Z}_p[x, y]$ by reducing A and B modulo p. In other words, if $\overline{A}$ and $\overline{B}$ are the residue classes of A and B in $\mathbb{Z}_p$, then $E : y^2 = x^3 + Ax + B$ can be reduced to $E_p : y^2 = x^3 + \overline{A}x + \overline{B}$ in $\mathbb{Z}_p[x, y]$.

If $p \neq 2, 3$ and $p \nmid 4A^3 + 27B^2$, then E_p will be an elliptic curve defined over $\mathbb{Z}_p$, since $4\overline{A}^3 + 27\overline{B}^2 \neq 0$ in $\mathbb{Z}_p$ and $\mathrm{Char}(\mathbb{Z}_p) \neq 2, 3$. In this case we say E has **good reduction** modulo p.

If the curve E_p has a multiple root ($4\overline{A}^3 + 27\overline{B}^2 = 0$), then E is said to have **bad reduction** modulo p.

Bad reduction comes in three different varieties;

1. If E_p has a triple root, then E has **additive reduction** modulo p.
2. If E_p has a double root and well-defined tangent line(s) everywhere, then E has **split multiplicative reduction** modulo p.
3. If E_p has a double root but does not have well-defined tangent lines everywhere, then E has **nonsplit multiplicative reduction** modulo p.

The necessity for making these distinctions will become clearer later.

5.3 Hecke L-Series and Hasse-Weil Zeta Functions

The goal of this section is to introduce two generalisations of the Riemann zeta function $\zeta(s)$, known as Hecke L-series and Hasse-Weil zeta functions. For every elliptic curve $E : y^2 = x^3 + Ax + B$ where $A, B \in \mathbb{Z}$, one can associate a Hecke L-series and a Hasse-Weil zeta function. The L-series, in particular, satisfy important functional relations and contain vital information about the structure of elliptic curves.

An L-series of E is constructed in a manner that preserves information about the cardinality $|E(\mathbb{F}_{p^n})|$ where $\mathbb{F}_{p^n}$ is an arbitrary finite field. From theorem 5.4, it is clear that the cardinality of $E(\mathbb{F}_{p^n})$ is completely determined by the cardinality of $E(\mathbb{F}_p)$. It is possible to construct a power series that respects this relation and has a simple closed form.

For any prime $p > 4$ not dividing $\Delta = 4A^3 + 27B^2$, define

$$\zeta_{p,E}(s) := \exp\left(\sum_{n=1}^{\infty} \frac{|E(\mathbb{F}_{p^n})|}{n} p^{-ns}\right). \tag{5.5}$$

Suppose that $a_p = p + 1 - |E(\mathbb{F}_p)|$ and $x^2 - a_p x + p = (x - \alpha)(x - \beta)$. Then using theorem 5.4, this becomes

$$\zeta_{p,E}(s) = \exp\left(\sum_{n=1}^{\infty} \frac{(p^n + 1 - \alpha^n - \beta^n)}{n} p^{-ns}\right).$$

Employing the series expansion

$$-\log(1 - x) = \sum_{n=1}^{\infty} \frac{x^n}{n},$$

this can be rewritten as

$$\zeta_{p,E}(s) = \exp\left(-\log(1 - p^{1-s}) - \log(1 - p^{-s}) + \log(1 - \alpha p^{-s}) + \log(1 - \beta p^{-s})\right).$$

Therefore,

$$\zeta_{p,E}(s) = \frac{(1 - \alpha p^{-s})(1 - \beta p^{-s})}{(1 - p^{-s})(1 - p^{1-s})} = \frac{1 - a_p p^{-s} + p^{1-2s}}{(1 - p^{-s})(1 - p^{1-s})}. \tag{5.6}$$

The product of this function can be taken over all suitable primes p to obtain a variant of the Riemann zeta function

$$\zeta_E^*(s) := \prod_{\substack{p \text{ prime} \\ p \nmid \Delta}} \zeta_{p,E}(s). \tag{5.7}$$

This is a naive definition for the Hasse-Weil zeta function of E. The complete definition will be outlined momentarily.

The term "zeta function" is indicative of the fact $\zeta_E^*(s)$ and $\zeta_{p,E}(s)$ share many properties with $\zeta(s)$. However, something interesting that sets them apart is the Riemann hypothesis. Obviously, this is still an unsolved problem for $\zeta(s)$. However, in 1973, Deligne [**8**] proved an analogous form of the hypothesis for generalised Hasse-Weil zeta functions. For elliptic curves this is relatively easy to prove.

Theorem 5.5. *Suppose* $E : y^2 = x^3 + Ax + B$ *is an elliptic curve where* $A, B \in \mathbb{Z}$*, then* $\zeta_{p,E}(s) = 0$ *implies* $\mathfrak{Re}(s) = \frac{1}{2}$.

Proof. Equation (5.6) guarantees that $\zeta_{p,E}(s) = 0$ if and only if $p^s = \alpha$ or $p^s = \beta$.

By Hasse's theorem, the polynomial $x^2 - a_p x + p$ is always non-negative over $\mathbb{R}$, thus, the roots α and β must be complex or equal. The polynomial also has real coefficients, so α and β must be conjugates. In particular, this implies $|\alpha| = |\beta| = p^{\mathfrak{Re}(s)}$. But $\alpha\beta = p$ by definition; therefore, $|\alpha||\beta| = p = p^{2\mathfrak{Re}(s)}$ and $\mathfrak{Re}(s) = \frac{1}{2}$. □

Returning now to the naive Hasse-Weil zeta function, equations (5.6) and (5.7) imply that the function may be rewritten as

$$\zeta_E^*(s) := \prod_{\substack{p \text{ prime} \\ p \nmid \Delta}} \frac{1 - a_p p^{-s} + p^{1-2s}}{(1 - p^{-s})(1 - p^{1-s})}.$$

The product

$$\prod_{\substack{p \text{ prime} \\ p \nmid \Delta}} \frac{1}{(1 - p^{-s})(1 - p^{1-s})}$$

is reasonably well understood due to its similarity to the Riemann zeta function. Thus, the non-trivial component of the product is described by the function

$$L_E^*(s) := \prod_{\substack{p \text{ prime} \\ p \nmid \Delta}} \left(\frac{1}{1 - a_p p^{-s} + p^{1-2s}} \right)$$

Here the term $1 - a_p p^{-s} + p^{1-2s}$ is inverted so that it may be expanded as a series. This function is an approximation to the Hecke L-series of E. In order to define the full Hecke L-series, it is necessary to include the primes that divide Δ in the product.

To accomplish this, the definition of a_p can be extended in the following way. Suppose $p \nmid \Delta$, then define

$$a_p := \begin{cases} 0 & \text{if } E \text{ has additve reduction modulo } p, \\ 1 & \text{if } E \text{ has split multiplicative reduction modulo } p, \\ -1 & \text{if } E \text{ has nonsplit multiplicative reduction modulo } p. \end{cases}$$

The full Hecke L-series of E is then defined as

$$L_E(s) := \prod_{\substack{p \text{ prime} \\ p|\Delta}} \left(\frac{1}{1 - a_p p^{-s}}\right) \prod_{\substack{p \text{ prime} \\ p\nmid\Delta}} \left(\frac{1}{1 - a_p p^{-s} + p^{1-2s}}\right).$$

The additional formulas for a_p are defined in such a way as to allow automorphic forms[1] to be constructed from $L_E(s)$. This is related to the Taniyama-Shimura conjecture, famously known for its use in Wiles' proof of Fermats last theorem.

What's more this extended definition of a_p implies that $L_E(s)$ satisfies a functional equation of the form

$$\Gamma(s)L_E(s) = C^{1-s}\Gamma(2-s)L_E(2-s)$$

where $\Gamma(s)$ is the gamma function and C is a constant dependent on E. This functional relation is analogous to one satisfied by $\zeta(s)$

$$\sqrt{\pi}\,\Gamma\left(\frac{s}{2}\right)\zeta(s) = \pi^s\,\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s).$$

Using Hecke L-series, it is not difficult to extend naive Hasse-Weil zeta functions to complete Hasse-Weil zeta functions. The Hasse-Weil zeta function is defined as

$$\zeta_E(s) := \frac{\zeta(s)\zeta(s-1)}{L_E(s)}.$$

[1] See [**1**], chapter 4 for more information on automorphic forms.

5.4 Complex Multiplication

Complex multiplication is the term given to a special property satisfied by certain elliptic curves. In section 2.1, the connection between elliptic curves and complex lattices was outlined. That connection was summarised by the **Alegbraic Law** on page 15.

Elliptic curves with complex multiplication correspond to lattices which have 'extra structure' or are more 'symmetric'. One way of interpreting how symmetric a lattice L is, is by considering its partial invariance under linear transformations. In particular, how L behaves under multiplication by elements of $\mathbb{C}$. Formally, the set of all relevant symmetries is

$$S_L := \{z \in \mathbb{C} \mid zL \subset L\}.$$

Clearly, lattices are abelian $\mathbb{Z}-$modules, so they must be closed under multiplication by an integer. In other words, the right coset nL is a subset of L for all $n \in \mathbb{Z}$. Therefore, $\mathbb{Z}$ must be a subset of S_L. For most lattices, this inclusion will be an equality. Thus, any extra structure a lattice L may have must correspond to the existence of a non-integer element in S_L i.e. $\alpha \in S_L \setminus \mathbb{Z}$. If such an element exists, L and the unique elliptic curve it gives rise to are said to have/exhibit **complex multiplication**.

Complex multiplication for elliptic curves can also be defined in terms of endomorphism rings. The following theorem highlights why.

Theorem 5.6. *Let E be the complex elliptic curve corresponding to the lattice L and let* $\mathrm{End}(E)$ *denote the ring of endomorphisms of E, then*

$$\mathrm{End}(E) \cong \{z \in \mathbb{C} \mid zL \subset L\}.$$

A proof can be found in [**30**], chapter 10. Multiplication-by-n endomorphisms in $\mathrm{End}(E)$ correspond uniquely to integers $n \in S_L$. If the set of multiplication endomorphisms is denoted $\hat{\mathbb{Z}}$ then E has complex multiplication if and only if $S_L \setminus \mathbb{Z}$ is non-empty, or equivalently, if and only if $\mathrm{End}(E) \setminus \hat{\mathbb{Z}}$ is non-empty.

Complex multiplication is ubiquitous in the theory of elliptic curves. Generally, when a conjecture is made the first course of action is to try and prove it for curves with complex multiplication because they are much easier to work with. The prototypical example of an elliptic curve with complex multiplication is

$$y^2 = x^3 + 4x.$$

This curve has a square lattice and an endomorphism ring containing the Gaussian integers, so it is very symmetric by any reasonable definition of symmetry.

5.5 Heights

In this section, we revise and extend what was discussed earlier in section 2.4 about heights. As before H denotes the multiplicative height and h denotes the logarithmic height of an elliptic curve $E(\mathbb{Q})$.

Property 2 of the logarithmic height guarantees that there exists a constant C_E, depending only on E, such that

$$4h(P) - h(2P) \leq C_E$$

for all $P \in E(\mathbb{Q})$.

This inequality can be extended to

$$|4h(P) - h(2P)| \leq C_E. \tag{5.8}$$

Preferably, we would want an equality between $4h(P)$ and $h(2P)$ rather than an inequality. This can be accommodated by altering the definition of h somewhat.

Definition 5.3. *The Néron–Tate height is the function $\hat{h} : E(\mathbb{Q}) \longrightarrow \mathbb{R}^+$ defined as the limit*

$$\hat{h}(P) := \frac{1}{2} \lim_{n\to\infty} \frac{1}{4^n} h(2^n P).$$

The Néron–Tate height not only satisfies $4\hat{h}(P) = \hat{h}(2P)$, but also $m^2\hat{h}(P) = \hat{h}(mP)$ for all $m \in \mathbb{Z}$. More generally, when adding elements of $E(\mathbb{Q})$ together, say P and Q, one has

$$\hat{h}(P+Q) + \hat{h}(P-Q) = 2\hat{h}(Q) + 2\hat{h}(P). \tag{5.9}$$

By taking appropriate limits, this formula is essentially a corollary of the fact that there exist $c_1, c_2 \in \mathbb{R}$ (only dependent on E) such that

$$2h(P) + 2h(Q) - c_1 \leq h(P+Q) + h(P-Q) \leq 2h(P) + 2h(Q) + c_2. \tag{5.10}$$

In order to prove (5.10), we shall make use of the following lemmas. These will not be proved here, however, full proofs can be found in [**30**], chapter 8.

Lemma 5.2. *Suppose $a_1, a_2, b_1, b_2 \in \mathbb{Z}$. Then*

$$\max\{|a_1|, |b_1|\} \max\{|a_2|, |b_2|\} \leq 2 \max\{|a_1a_2|, |b_1b_2|, |a_1b_2 + b_1a_2|\}.$$

Lemma 5.3. *Suppose $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ and $\gcd(a_1, b_1) = \gcd(a_2, b_2) = 1$. Then*

$$\gcd(a_1a_2, b_1b_2, a_1b_2 + a_2b_1) = 1.$$

Then, the proof of (5.10) is as follows.

Proof. Suppose $E : y^2 = x^3 + Ax + B$ is an elliptic curve and $A, B \in \mathbb{Z}$. For any $P, Q \in E(\mathbb{Q})$ write

$$P = \left(\frac{p_1}{q_1}, y_1\right), \qquad Q = \left(\frac{p_2}{q_2}, y_2\right),$$
$$P + Q = \left(\frac{p_3}{q_3}, y_3\right), \quad P - Q = \left(\frac{p_4}{q_4}, y_4\right),$$

where $\gcd(p_i, q_i) = 1$ for all $i \in \{1, 2, 3, 4\}$.

The addition formulas for elliptic curves in section 2.2 can be used to determine that

$$\frac{p_3}{q_3} + \frac{p_4}{q_4} = \frac{r_1}{r_3}, \quad \frac{p_3 p_4}{q_3 q_4} = \frac{r_2}{r_3}$$

where

$$r_1 = 2(p_1 q_2 + q_1 p_2)(A q_1 q_2 + p_1 p_2) + 4B q_1^2 q_2^2,$$
$$r_2 = (p_1 p_2 - A q_1 q_2)^2 - 4B(p_1 q_2 + q_1 p_2) q_1 q_2,$$
$$r_3 = (p_1 q_2 - q_1 p_2)^2.$$

Applying lemma 5.3 to p_3, p_4, q_3, q_4 it follows that $\gcd(p_3 p_4, q_3 q_4, p_3 q_4 + p_4 q_3) = 1$, and consequently that there exist $x, y, z \in \mathbb{Z}$ such that

$$p_3 p_4 x + q_3 q_4 y + (p_3 q_4 + p_4 q_3) z = 1.$$

Multiplying across this equation by r_3 and using the above formulas, it can be rewritten as

$$r_2(q_3 q_4)x + r_3(q_3 q_4)y + r_1(q_3 q_4)z = r_3.$$

This implies $(q_3q_4)|r_3$, from which it follows that $(p_3p_4)|r_2$ and $(p_3q_4 + p_4q_3)|r_1$ as well. Therefore,

$$|p_3q_4 + p_4q_3| \leq |r_1|, \quad |p_3p_4| \leq |r_2|, \quad |q_3q_4| \leq |r_3|.$$

Now, by definition of the multiplicative height H,

$$H(P + Q)H(P - Q) = \max\{|p_3|, |q_3|\} \max\{|p_4|, |q_4|\}.$$

Which by lemma 5.2 and the above discussion satisfies

$$H(P + Q)H(P - Q) \leq 2 \max\{|r_1|, |r_2|, |r_3|\}. \tag{5.11}$$

The next step in the proof involves finding upper bounds for $|r_1|, |r_2|$ and $|r_3|$.

Clearly, $H(P) = \max\{|p_1|, |q_1|\}$ and $H(Q) = \max\{|p_2|, |q_2|\}$. Consequently,

$$\begin{aligned} |r_1| &= |2(p_1q_2 + q_1p_2)(Aq_1q_2 + p_1p_2) + 4Bq_1^2q_2^2| \\ &\leq 4(1 + |A| + |B|)H(P)^2H(Q)^2, \end{aligned}$$

$$\begin{aligned} |r_2| &= |(p_1p_2 - Aq_1q_2)^2 - 4B(p_1q_2 + q_1p_2)q_1q_2| \\ &\leq ((1 - A)^2 + 8B)H(P)^2H(Q)^2, \end{aligned}$$

$$|r_3| = |(p_1q_2 - q_1p_2)|^2 \leq 4H(P)^2H(Q)^2.$$

Choose $C = \max\{4(1 + |A| + |B|), 4, ((1 - A)^2 + 8B)\}$. Then $|r_i| \leq CH(P)^2H(Q)^2$ for all $i \in \{1, 2, 3\}$. Hence,

$$H(P + Q)H(P - Q) \leq 2CH(P)^2H(Q)^2$$

by (5.11).

Taking logarithms yields

$$h(P + Q) + h(P - Q) \leq 2h(P) + 2h(Q) + \log(2C)$$

which proves the upper bound in (5.10).

To prove the lower bound, simply make the change of variables $P \longrightarrow P + Q$ and $Q \longrightarrow P - Q$ to get

$$h(2P) + h(2Q) \leq 2h(P + Q) + 2h(P - Q) + \log(2C).$$

Equation (5.8) implies

$$4h(P) + 4h(Q) - 2C_E \leq h(2P) + h(2Q) \leq 2h(P+Q) + 2h(P-Q) + \log(2C).$$

Therefore,

$$2h(P) + 2h(Q) - C_E - \log(\sqrt{2C}) \leq h(P+Q) + h(P-Q)$$

which completes the proof of (5.10) and in turn (5.9).

□

Part of the reason the Néron–Tate height is so important is because equation (5.9) can be used to induce a bilinear mapping on $E(\mathbb{Q})$.

Definition 5.4. *The **height pairing** of E is the map $\langle *, * \rangle : E(\mathbb{Q}) \times E(\mathbb{Q}) \longrightarrow \mathbb{R}$ defined by*

$$\langle P, Q \rangle = \hat{h}(P+Q) - \hat{h}(P) - \hat{h}(Q).$$

Clearly $\langle *, * \rangle$ is symmetric so proving bilinearity is equivalent to proving linearity in one variable, i.e.

$$\langle P_1 + P_2, Q \rangle = \langle P_1, Q \rangle + \langle P_2, Q \rangle.$$

Proof. Repeated use of (5.9) implies

$$\begin{gathered}
\hat{h}(P_1+P_2+Q) + \hat{h}(P_1+P_2-Q) - 2\hat{h}(P_1+P_2) - 2\hat{h}(Q) = 0, \\
\hat{h}(P_1+P_2+Q) + \hat{h}(P_1-P_2+Q) = 2\hat{h}(P_1+Q) + 2\hat{h}(P_2), \\
-\hat{h}(P_1+P_2-Q) - \hat{h}(P_1-P_2+Q) = -2\hat{h}(P_2-Q) - 2\hat{h}(P_1), \\
0 = 2\hat{h}(P_2-Q) + 2\hat{h}(P_2+Q) - 4\hat{h}(P_2) - 2\hat{h}(Q).
\end{gathered}$$

Adding these four equations together yields

$$\begin{gathered}
2\left(\hat{h}(P_1+P_2+Q) - \hat{h}(P_1+P_2) - \hat{h}(Q)\right) = \\
2\left(\hat{h}(P_1+Q) - \hat{h}(P_1) - \hat{h}(Q) + \hat{h}(P_2+Q) - \hat{h}(P_2) - \hat{h}(Q)\right).
\end{gathered}$$

Therefore,

$$2\langle P_1 + P_2, Q \rangle = 2\langle P_1, Q \rangle + 2\langle P_2, Q \rangle.$$

Dividing by 2 gives the desired result.

□

If $P_1, P_2, \ldots, P_n$ are points on $E(\mathbb{Q})$, the height pairing can be used to determine whether they are independent. In other words, whether or not there exist integers $k_1, k_2, \ldots, k_n \in \mathbb{Z}$, not all zero, such that

$$k_1P_1 + k_2P_2 + \cdots + k_nP_n = \infty.$$

Consider the matrix A whose ij-entry, $A_{i,j}$, is $\langle P_i, P_j \rangle$. If the determinant of A is non-zero, then $P_1, P_2, \ldots, P_n$ are independent. This follows from the fact that non-singular matrices must have linearly independent rows and columns.

If $P_1, P_2, \ldots, P_n$ is a complete list of independent, infinite order generators for $E(\mathbb{Q})$, then the determinant $R_E = \det(A)$ is known as the **elliptic regulator** of E. In the case where $E(\mathbb{Q})$ has no elements of infinite order, define $R_E := 1$.

5.6 Modular Curves

This section provides a very brief introduction to **modular curves**.

Given a complex elliptic curve E, the Algebraic Law guarantees that E corresponds uniquely to a lattice L in $\mathbb{C}$. However, this does *not* imply that E corresponds uniquely to a basis for L. It may happen that two distinct bases $\{\omega_1, \omega_2\}$ and $\{\omega_1', \omega_2'\}$ generate the same lattice, and thus, correspond to the same elliptic curve. For example, if a lattice L is generated by $\{\omega_1, \omega_2\}$, then it is also generated by $\{\omega_1, \omega_1 + \omega_2\}$. The following theorem outlines the necessary and sufficient conditions for two bases to generate the same lattice.

Theorem 5.7. *Suppose $\omega_1, \omega_2 \in \mathbb{C}$ are linearly independent over $\mathbb{R}$ and $\omega_1', \omega_2' \in \mathbb{C}$ are also linearly independent over $\mathbb{R}$. Then $\{\omega_1, \omega_2\}$ and $\{\omega_1', \omega_2'\}$ generate the same lattice in $\mathbb{C}$ if and only if*

$$\begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \tag{5.12}$$

for some integer matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with determinant $ad - bc = \pm 1$.

It is possible to show (though it will not be done here) that scaling bases

$$\{\omega_1, \omega_2\} \longrightarrow \{\lambda\omega_1, \lambda\omega_2\},\ \lambda \in \mathbb{C} \setminus \{0\}$$

will scale the corresponding elliptic curve accordingly:

$$E : y^2 = x^3 + Ax + B \longrightarrow E_\lambda : y^2 = x^3 + \lambda^4 Ax + \lambda^6 B.$$

For all intents and purposes, E and E_λ can be thought of as the same elliptic curve (consider the isomorphism introduced in lemma 2.1). Thus, by letting $\lambda = \omega_2^{-1}$, it suffices to work with bases of the form $\{\frac{\omega_1}{\omega_2}, 1\}$. By relabelling ω_1 and ω_2 if necessary, it can always be assumed $\Im(\frac{\omega_1}{\omega_2}) > 0$.

For notational convenience, $\frac{\omega_1}{\omega_2}$ is denoted τ. The upper half plane $\{z \in \mathbb{C} \mid \Im(z) > 0\}$ is denoted $\mathbb{H}$. Thus, the above assumption says that $\tau \in \mathbb{H}$.

It follows from (5.12) that we have a certain degree of freedom for the choice of τ. More specifically, if (5.12) holds, then $\{\omega_1, \omega_2\}$, $\{\omega_1', \omega_2'\}$ give rise to the same lattice and there exist integers a, b, c, d with $ad - bc = \pm 1$ such that

$$\tau' := \frac{\omega_1'}{\omega_2'} = \frac{a\omega_1 + b\omega_2}{c\omega_1 + d\omega_2} = \frac{a\tau + b}{c\tau + d}.$$

Therefore, to guarantee τ depends only on a choice of lattice, not basis, τ and $\tau' = \left(\frac{a\tau + b}{c\tau + d}\right)$ should be regarded as equivalent.

Establishing an equivalence relation is not difficult, but does involve a couple of caveats. Since our concern is only with elements of $\mathbb{H}$, it is necessary that τ' be an element of $\mathbb{H}$ whenever τ is, and vice versa. By making the stronger assumption that $ad - bc = 1$, this property will hold. As an example of why the assumption $ad - bc = -1$ is no longer applicable, consider $a = -1$, $d = 1$ and $b = c = 0$. Then $ad - bc = -1$ and

$$\tau' = \left(\frac{a\tau + b}{c\tau + d}\right) = -\tau.$$

Therefore, $\Im(\tau) = -\Im(\tau')$, so τ and τ' cannot both lie in $\mathbb{H}$.

The assumption that $ad - bc = 1$ can be rephrased as saying we only consider matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$$

where $\mathrm{SL}_2(\mathbb{Z})$ is the matrix group defined as

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

The second caveat comes from the fact

$$\tau' = \frac{a\tau + b}{c\tau + d} = \frac{-a\tau - b}{-c\tau - d} \text{ for all } \tau \in \mathbb{H}.$$

Thus, there is a degeneracy when it comes to the choice of matrix in $\mathrm{SL}_2(\mathbb{Z})$, in so far as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ and } \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$$

give rise to the same τ'.

To overcome this problem, the quotient group

$$\mathrm{PSL}_2(\mathbb{Z}) := \mathrm{SL}_2(\mathbb{Z})/\{I, -I\}$$

is used. This quotient group identifies every matrix $M \in \mathrm{SL}_2(\mathbb{Z})$ with its additive inverse $-M$. Usually $\mathrm{PSL}_2(\mathbb{Z})$ is referred to as the **modular group** and denoted[2] Γ.

The modular group induces a left group action on the upper half plane $\mathbb{H}$ via the map

$$\rho : \Gamma \times \mathbb{H} \longrightarrow \mathbb{H},\ \rho(M, \tau) = \frac{a\tau + b}{c\tau + d}$$

where

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

[2]Not to be confused with the absolute Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

The orbits of the group action ρ admit a partition of the upper half plane $\mathbb{H}$. In other words, given any two elements of $\mathbb{H}$, say τ and τ', the relation

$$\tau \sim \tau' \text{ if and only if } \tau' = \rho(M, \tau) \text{ for some } M \in \Gamma,$$

is an equivalence relation.

Topologically, the set of all equivalence classes $\mathbb{H}/\Gamma$ can be realised as a Riemann surface.[3] This is an example of a **non-compact modular curve**. More generally, a non-compact modular curve is a Riemann surface that can be constructed as a quotient space $\mathbb{H}/\Gamma^*$, where Γ^* must be a subgroup of Γ of finite index. A **compact modular curve** is a compactification of $\mathbb{H}/\Gamma^*$ obtained by adding finitely many points.

An important property that some modular curves have is that they can be used to parametrise elliptic curves. An elliptic curve that can be parametrised in this way is known as a **modular elliptic curve**. These special elliptic curves have huge historical significance in mathematics, as shall be seen in the next and final chapter.

[3] A Riemann surface is a connected one dimensional complex manifold.

Chapter 6

The Conjecture

In 1965, two years after the publication of *Notes on elliptic curves I* [**26**], Birch and Swinnerton-Dyer completed their subsequent paper *Notes on elliptic curves II* [**27**]. Having developed an algorithm for bounding the rank of elliptic curves, they were able to substantiate the following conjecture for their second paper.

The Birch and Swinnerton-Dyer conjecture. *Let r be the rank of a rational elliptic curve E, then $\zeta_E(s)$ has a pole of order $r+1$ at $s=1$.*

Equivalently,

The Birch and Swinnerton-Dyer conjecture. *Let r be the rank of a rational elliptic curve E, then $L_E(s)$ has a zero of order r at $s=1$.*

This was a remarkable conjecture, as it was not even known at that time whether $L_E(s)$ could be defined at $s=1$ for an arbitrary elliptic curve. In 1953, Deuring [**9**] proved that if E had complex multiplication, then $L_E(s)$ could be analytically extended to the entire complex plane. However, only after Wiles et al. proved the Taniyama-Shimura conjecture was it established that $L_E(s)$ could be analytically continued for any elliptic curve E.

This conjecture arose from studying the *formal* definition of $L_E(s)$ at $s=1$,

$$L_E(1) := \prod_{\substack{p \text{ prime} \\ p|\Delta}} \left(\frac{1}{1-a_p p^{-1}}\right) \prod_{\substack{p \text{ prime} \\ p\nmid\Delta}} \left(\frac{1}{1-a_p p^{-1}+p^{-1}}\right).$$

This can also be written as

$$L_E(1) := \prod_{\substack{p \text{ prime} \\ p|\Delta}} \left(\frac{1}{1 - a_p p^{-1}} \right) \prod_{\substack{p \text{ prime} \\ p \nmid \Delta}} \left(\frac{p}{|E(\mathbb{F}_p)|} \right).$$

For earlier work in the 1960s, Birch and Swinnerton-Dyer employed the use of the EDSAC-2 computer system to calculate $|E(\mathbb{F}_p)|$ for large quantities of primes. Using this computer system alongside the algorithm they developed in [**26**], they were also able to bound, and in many cases explicitly determine, the rank r of any elliptic curve E. Comparing both sets of data, they came to suspect that

$$f(x) := \prod_{\substack{p \text{ prime} \\ p \nmid \Delta \\ p \leq x}} \left(\frac{p}{|E(\mathbb{F}_p)|} \right) \longrightarrow C \log(x)^{-r} \text{ as } x \longrightarrow \infty \tag{6.1}$$

for some non-zero constant C, depending only on E. In other words, that the function $f(x)$, which is an approximation to $L_E(1)$, converges at a rate proportional to $\log(x)^{-r}$. The Birch and Swinnerton-Dyer conjecture is a weaker form of this assertion.

The majority of the second paper revolves around determining $L_E(1)$ for rational elliptic curves of the form

$$E_D : y^2 = x^3 - Dx$$

where D is a non-zero integer. Elliptic curves of this form are particularly easy to deal with because their corresponding lattices are square. In particular, this means they must admit complex multiplication (due to the invariance of a square lattice under a 90° rotation), so Deurings result mentioned earlier applies.

Birch and Swinnerton-Dyer were able to partially validate their conjecture for the curves E_D, in so far as they obtained a great deal of evidence to support the proposition

$$L_D(1) = 0 \text{ if and only if } r > 0 \tag{6.2}$$

where as before, r denotes the rank of E_D. For notational convenience, the subscript D will often be used instead of E_D from now on. This proposition is of course implied by the Birch and Swinnerton-Dyer conjecture.

This hypothesis is supported by the extensive tables Birch and Swinnerton-Dyer included in their second paper. These tables include the ranks of hundreds of elliptic curves. The method they originally developed in [**26**] for calculating/bounding ranks proved too inefficient for large values of D, so they refined it to suit the curves E_D specifically.

In the remaining case, when $r = 0$, Birch and Swinnerton-Dyer were unable to interpret the value of $L_D(1)$ from computation alone. They were even unable to predict its sign. As a result, they sought advice from Davenport, who suggested that $L_D(1)$ should be expressible in a specific closed form. Along with suggestions from Kneser that this closed form should contain an integer, they were able to deduce and prove that

$$L_D(1) = \begin{cases} \dfrac{C'\sigma(D)}{\sqrt[4]{D}} & \text{if } D > 0, \\ \dfrac{C'\sigma(D)}{\sqrt[4]{-4D}} & \text{if } D < 0. \end{cases}$$

where C' is a non-zero constant and $\sigma(D)$ is an integer[1] depending on E_D. On the following page is a table of values from [**27**] displaying D, the rank of E_D and the corresponding value of $\sigma(D)$. Birch and Swinnerton-Dyer denote the rank of E_D by g and shorten $\sigma(D)$ to σ.

[1] Technically, $\sigma(D)$ is only an integer when $|D| > 1$ and D is not divisible by 4 or a fourth power.

D	D			$-D$			$2D$			$-2D$		
	σ	g	$\lambda.\lambda_1$	σ	g	$\lambda.\lambda_1$	σ	g	$\lambda.\lambda_1$	σ	g	$\lambda.\lambda_1$
1	¼	0	1.1	½	0	0.2	0	1	2.1	1	0	1.1
3	1	0	1.1	0	1	1.2	0	1	2.1	2	0	1.1
5	0	1	2.1	0	1	1.2	0	1	2.1	2	0	1.1
7	0	1	1.2	2	0	1.1	0	1	2.1	0	2	2.2
9	1	0	2.0	0	1	0.3	2	0	1.1	0	1	2.1
11	1	0	1.1	2	0	1.1	0	1	2.1	2	0	1.1
13	2	0	1.1	0	1	1.2	0	1	2.1	2	0	1.1
15	0	1	1.2	0	1	1.2	0	1	2.1	4	0	1.1
17	0	2	2.2	4*	0	1.3	0	1	3.2	0	2	2.2
19	1	0	1.1	0	1	1.2	0	1	2.1	2	0	1.1
21	0	1	2.1	0	1	1.2	0	1	2.1	4	0	1.1
23	0	1	1.2	2	0	1.1	0	1	2.1	0	2	2.2
25	0	1	2.1	2	0	0.2	0	1	2.1	2	0	1.1
27	1	0	1.1	2	0	1.1	0	1	2.1	2	0	1.1
29	2	0	1.1	0	1	1.2	0	1	2.1	2	0	1.1
31	0	1	1.2	0	1	1.2	0	1	2.1	8*	0	2.2
33	4	0	1.1	0	2	1.3	0	1	2.1	0	2	2.2
35	2	0	1.1	0	1	1.2	0	1	2.1	4	0	1.1
37	0	1	2.1	0	1	1.2	0	1	2.1	2	0	1.1
39	0	1	1.2	0	2	2.2	0	1	2.1	4	0	1.1
41	0	1	2.1	4*	0	1.3	0	3	3.2	8*	0	2.2
43	1	0	1.1	2	0	1.1	0	1	2.1	2	0	1.1
45	0	1	1.2	4	0	1.1	0	2	3.1	0	1	1.2
47	0	1	1.2	0	1	1.2	0	1	2.1	0	2	2.2
49	0	1	2.1	0	1	0.3	4*	0	3.1	0	1	1.2
51	2	0	1.1	0	1	1.2	0	1	2.1	4	0	1.1
53	0	1	2.1	0	1	1.2	0	1	2.1	2	0	1.1
55	0	1	1.2	0	2	2.2	0	1	2.1	4	0	1.1
57	0	1	2.1	8*	0	1.3	0	1	2.1	0	2	2.2
59	1	0	1.1	2	0	1.1	0	1	2.1	2	0	1.1
61	2	0	1.1	0	1	1.2	0	1	2.1	2	0	1.1
63	4	0	1.1	0	2	1.3	4	0	1.1	0	1	2.1
65	0	2	2.2	0	2	1.3	0	1	2.1	4	0	1.1
67	1	0	1.1	0	1	1.2	0	1	2.1	2	0	1.1
69	0	1	2.1	0	1	1.2	0	1	2.1	0	2	2.2
71	0	1	1.2	2	0	1.1	0	1	2.1	8*		2.2
73	0	1	2.1	0	2	1.3	0	1	3.2	8*		2.2
75	2	0	1.1	4	0	1.1	0	1	2.1	0	2	2.2
77	0	2	2.2	0	1	1.2	0	1	2.1	0	2	2.2
79	0	1	1.2	0	1	1.2	0	1	2.1	0	2	2.2

This figure includes 160 different values for D, σ and g. Studying the table, one notes that σ is essentially always a positive integer whenever $g = 0$. This trend continues for every value of D that Birch and Swinnerton-Dyer consider in their tables. Although it does not prove it, the tendency for this relationship to occur does help substantiate (6.2), and in turn, the Birch and Swinnerton-Dyer conjecture.

6.1 Consequences

In this penultimate section, several consequences of the Birch and Swinnerton-Dyer conjecture (under the assumption it is true) are outlined.

One of the most famous corollaries of the Birch and Swinnerton-Dyer conjecture involves the classical concept of a *congruent number*.

Definition 6.1. *A natural number $n \in \mathbb{N}$ is said to be **congruent** if is the area of some rational right angle triangle i.e. if there exists $a, b, c \in \mathbb{Q}$ such that*

$$a^2 + b^2 = c^2 \text{ and } \frac{1}{2}ab = n.$$

A theorem due to Jerrold Tunnell, see [**29**], says that if $n \in \mathbb{N}$ is a congruent number, then the cardinalities of the following sets are equal.

$$|\{(x, y, z) \in \mathbb{Z}^3 \mid 2x^2 + y^2 + 8z^2 = n, z \text{ even}\}| =$$
$$|\{(x, y, z) \in \mathbb{Z}^3 \mid 2x^2 + y^2 + 8z^2 = n, z \text{ odd}\}|$$

He also showed that the converse holds if the Birch and Swinnerton-Dyer conjecture is true. In other words, the equality above is strong enough to imply that n is a congruent number.

Another interesting theorem contingent on the conjecture involves a property of rational dynamical systems. Michael Stoll, see [**25**], proved the following.

Theorem 6.1. *If the Birch and Swinnerton-Dyer conjecture is true, then there does not exist $a, b \in \mathbb{Q}$ such that the sequence defined by*

$$x_1 = a, x_{n+1} = x_n^2 + b$$

satisfies $x_{n+6} = x_n$, for all large enough $n \in \mathbb{N}$, unless it also satisfies $x_{n+2} = x_n$ or $x_{n+3} = x_n$. In other words, periodic 6-cycles of this sequence that are not also 2 or 3-cycles cannot exist.

A result that has already been superseded, but is still worth mentioning assumes the truth of the Birch and Swinnerton-Dyer conjecture and a stronger form of the Riemann hypothesis, see [**15**].

Theorem 6.2. *If the Birch and Swinnerton-Dyer conjecture and the generalised Riemann hypothesis are true, then the average rank of a rational elliptic curve is less than* 2*.*

We shall see in the next section that in work by Manjul Bhargava and Arul Shankar a stronger, unconditional version of this result has been proved.

Other applications of the conjecture are computational in nature. For example, it is well-known that the truth of the conjecture would imply the existence of an effective algorithm for calculating the rank of any rational elliptic curve E. This, in particular, would also imply there exists an algorithm for computing $E(\mathbb{Q})$, via Mordell's theorem and Mazur's theorem.

Being able to compute the rank of an elliptic curve is essential for many areas of algebraic number theory, particularly in relation to Gauss' class number problem for quadratic field extensions, see [**10**]. It is currently unknown whether elliptic curves can have arbitrarily large ranks. To date the highest known rank is $r = 20$, which was discover by Elkies and Zev Klagsbrun. It is hoped that the Birch and Swinnerton-Dyer conjecture may provide a partial answer to this unsolved problem.

6.2 Progress

Over the past several decades, the Birch and Swinnerton-Dyer conjecture has gone on to become one of the most (in)famous problems in mathematics. So much so that it was chosen by the Clay Mathematics Institute as one of only seven millennium prize problems, for which there is a one million dollar reward for solving.

Although the conjecture has cemented itself as one of the most challenging problems in modern mathematics, some progress has still been made.

First and foremost, the conjecture itself has been refined and strengthened many times since its inception. The original form of the conjecture makes no reference to the leading coefficient of $L_E(s)$ at $s = 1$, only the proposed order. However, the modern formulation of the conjecture includes the suspected form of

$$\lim_{s\to 1}\left(\frac{L_E(s)}{(s-1)^r}\right).$$

Explicitly, one has

The Strong Birch and Swinnerton-Dyer conjecture. *Let r be the rank of a rational elliptic curve E, then*

$$L_E(s) = \left(\frac{|Ш_E| R_E \Omega_\infty}{|\tau_E|^2} \prod_{p|\Delta} \Omega_p\right)(s-1)^r + O((s-1)^{r+1})$$

where $O((s-1)^{r+1})$ denotes terms of higher order.

An explanation of some of the terms used above is required. Firstly, $|Ш_E|$ denotes the order of the Tate-Shafarevich group defined in section 3.2 and R_E denotes the elliptic regulator defined at the end of section 5.5. The subgroup of $E(\mathbb{Q})$ consisting of all points of finite order is denoted τ_E and is known as the torsion subgroup of $E(\mathbb{Q})$. The term $|\tau_E|^2$ denotes the square of this group's order. As usual, Δ denotes the discriminant of E and p denotes a prime. The factor Ω_∞ is defined below. A full definition of the factors $\{\Omega_p\}$ is not included, but can be found in [**28**].

The Ω_∞ term is simple to define. Consider the rational elliptic curve E with discriminant Δ, and suppose the associated lattice L is generated by $\{\omega_1, \omega_2\}$. Because E has coefficients in $\mathbb{Q}$ and thus, in $\mathbb{R}$, it can be assumed that either ω_1 or ω_2, but not both, is a real number, see [**30**], chapter 9. By relabelling if necessary, assume that $\omega_2 \in \mathbb{R}$. Then Ω_∞ can be defined as

$$\Omega_\infty := \begin{cases} \omega_2 & \text{if } \Delta < 0, \\ 2\omega_2 & \text{if } \Delta > 0. \end{cases}$$

On the other hand, the factors Ω_p are much harder to define. They are known as *Tamagawa numbers*. These factors are the orders of certain quotient groups of $E(\mathbb{Q}_p)$. If E has good reduction at p, then $\Omega_p = 1$. This is not necessarily true for primes at which E has bad reduction. Thus, Ω_p can be thought of as giving a rough measure to how 'close' E is to having good reduction at p, as per section 5.2.

Concerning the appearance of the Tate-Shafarevich group in the conjectures formulation, John Tate famously remarked that "*This remarkable conjecture relates the behaviour of a function L at a point where it is not at present known to be defined to the order of a group* Ш *which is not known to be finite*"

One of the first breakthroughs in resolving the conjecture was accomplished in 1977 by John Coates and Andrew Wiles (the same Andrew Wiles that proved Fermats last theorem). Wiles was a doctoral student of Coates at the time. They proved the following in [**7**].

Coates-Wiles. *Let E be a rational elliptic curves which exhibits complex multiplication and suppose $L_E(1) \neq 0$, then E has rank $r = 0$.*

An important case where the above theorem applies is elliptic curves of the form $E_D : y^2 = x^3 - Dx$, $D \in \mathbb{Z} \setminus \{0\}$. This is the same family of curves that Birch and Swinnerton-Dyer considered in [**27**].

The proof of this theorem involves rewriting $L_E(1)$ as a product and showing that the existence of a point of infinite order in $E(\mathbb{Q})$ forces one of the factors to be zero, contradicting the assumption that $L_E(1)$ is non-zero. The techniques used to establish this contradiction involve results from Galois theory and algebraic number theory, in particular, the theory of cyclotomic fields.

Six years later, in 1983, Ralph Greenberg proved a partial converse to the Coates-Wiles theorem. In [**11**], he proved

Greenberg. *If a rational elliptic curve E exhibits complex multiplication and $L_E(s)$ has an odd order zero at $s = 1$, then either;*

1. *The rank of E is non-zero*

   or,

2. *the group $\bigcup_{n=1}^{\infty} Ш_E[p^n]$ is infinite for infinitely many primes p.*

What makes this theorem particularly important is that the second condition is thought to be highly unlikely. Therefore, if the second condition were to be disproved, there would be strong evidence in favour of the Birch and Swinnerton-Dyer conjecture.

Several years after Greenberg's result, Benedict Gross and Don Zagier [**12**] established a special case of the conjecture for elliptic curves that do not necessarily exhibit complex multiplication. Instead, they considered modular elliptic curves, see section 5.6

Gross-Zagier. *Suppose E is a modular elliptic curve defined over $\mathbb{Q}$ and $L_E(s)$ has a zero of order one at $s = 1$, then the rank of $E(\mathbb{Q})$ is non-zero, i.e. $E(\mathbb{Q})$ is infinite.*

To prove this theorem Gross and Zagier utilised the idea of a *Heegner point*. A Heegner point is a special type of point on modular curves that arises from quadratic number fields. They were originally defined by Birch. Heegner points are used in [**12**] to explicitly construct infinite order rational points on $E(\mathbb{Q})$.

The next major breakthrough was established by Karl Rubin in 1987, the year after Gross and Zagier's result. He proved the following theorems in [**20**].

Rubin 1. *Suppose E is a rational elliptic curve which exhibits complex multiplication and $L_E(1) \neq 0$, then $Ш_E$ is finite.*

Rubin 2. *Suppose E is a rational elliptic curve which exhibits complex multiplication and has rank $r \geq 2$, then $L_E(s)$ has a zero of order 2 or more at $s = 1$.*

The proofs of both theorems rely heavily on the work of Coates and Wiles. The second theorem is also dependent on Gross and Zagier's result.

The paper [**20**] is particularly famous in the history of mathematics as it provided the first example of a finite Tate-Shafarevich group, which is shown below.

The elliptic curve $E_1 : y^2 = x^3 - x$ has a trivial Tate-Shafarevich group i.e.

$$Ш_{E_1} \cong \{e\}.$$

An example of an elliptic curve that has a non-trivial, finite Tate-Shafarevich group is $E_{-17} : y^2 = x^3 + 17x$. The order of this group is 4. In fact

$$Ш_{E_{-17}} \cong \mathbb{Z}_2 \times \mathbb{Z}_2.$$

As Rubin highlights in his paper, his second theorem along with the results of Coates, Wiles, Gross and Zagier is strong enough to imply

Rubin et al. *Suppose E is a rational elliptic curve which exhibits complex multiplication, then the order of $L_E(s)$ at $s = 1$, denoted ρ, is equal to the rank of $E(\mathbb{Q})$ when $\rho \leq 1$.*

In 1989, a stronger form of this theorem was proven by Victor Kolyvagin, see [**17**].

Kolyvagin. *Suppose E is a modular elliptic curve defined over $\mathbb{Q}$, then the order of $L_E(s)$ at $s = 1$, denoted ρ, is equal to the rank of $E(\mathbb{Q})$ when $\rho \leq 1$. Furthermore, under the hypothesis of this theorem, $Ш_E$ will also be finite when $\rho \leq 1$.*

Kolyvagin's result was subsequently generalised a decade later when Taylor et al. [**3**] proved that all rational elliptic curves are modular. In other words,

Kolyvagin et al. *The Birch and Swinnerton-Dyer conjecture is true in the cases where $L_E(s)$ has a zero of order 0 or order 1.*

This is a strong result in support of the Birch and Swinnerton-Dyer conjecture.

The final, incredible result that will be mentioned is due to Manjul Bhargava and Arul Shankar, [**2**]. In 2015, they were able to prove that the 'average' rank of a rational elliptic curve E is bounded above by $\frac{7}{6}$. From this, they were also able to conclude that a positive proportion (over 66%) of rational elliptic curves have $L_E(1) \neq 0$ i.e. $L_E(s)$ has a zero of order 0 at $s = 1$. Employing Kolyvagin's result, this of course implies

Kolyvagin, Bhargava, Shankar et al. *The majority of rational elliptic curves satisfy the Birch and Swinnerton-Dyer conjecture.*

As for the future of the Birch and Swinnerton-Dyer conjecture, there is still a ways to go. Many special cases have been established, but as it stands elliptic curves with ranks greater than one are completely elusive.

Appendix A

Projective Geometry

In sections 2.2 and 2.3, the necessity of adding an identity element, ∞, to elliptic curve groups was highlighted. In this appendix, this idea will be made rigorous.

Consider a rational elliptic curve $E : y^2 = x^3 + Ax + B$. By choosing an appropriate denominator, any rational solution to this elliptic curve can be written in the form

$$\left(\frac{a}{c}, \frac{b}{c}\right) \in E(\mathbb{Q}) = \{(x, y) \in \mathbb{Q} \times \mathbb{Q} \mid y^2 = x^3 + Ax + B\}.$$

where $a, b, c \in \mathbb{Z}$.

This, in particular, implies that the triple (a, b, c) is a solution to the equation

$$\Gamma(x, y, z) := y^2z - x^3 - Axz^2 - Bz^3 = 0.$$

The polynomial Γ is known as the **homogenisation** of E with respect to the variable z.

Definition A.1. *Let $\mathbb{F}$ be an arbitrary field. A polynomial $f(x_1, x_2, \ldots, x_n) \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ is **homogeneous** if every monomial term of f has the same degree.*

For any $\lambda \in \mathbb{F}$, a homogeneous polynomial f will satisfy

$$f(\lambda x_1, \lambda x_2, \ldots, \lambda x_n) = \lambda^{\deg(f)} f(x_1, x_2, \ldots, x_n).$$

Notably, if $\lambda \in \mathbb{F} \setminus \{0\}$, this property implies $(x_1, x_2, \ldots, x_n)$ is a solution to $f = 0$ if and only $(\lambda x_1, \lambda x_2, \ldots, \lambda x_n)$ is too.

In relation to Γ, this makes sense since

$$\left(\frac{\lambda a}{\lambda c}, \frac{\lambda b}{\lambda c}\right) = \left(\frac{a}{c}, \frac{b}{c}\right) \in E(\mathbb{Q}), \text{ for all } \lambda \in \mathbb{Z}/\{0\}.$$

Thus, (a, b, c) is a solution to Γ whenever $(\lambda a, \lambda b, \lambda c)$ is.

Now, because

$$\left(\frac{a}{c}, \frac{b}{c}\right) \text{ and } \left(\frac{\lambda a}{\lambda c}, \frac{\lambda b}{\lambda c}\right)$$

represent the same solution of $E : y^2 = x^3 + Ax + B$, we would want (a, b, c) and $(\lambda a, \lambda b, \lambda c)$ to represent the 'same' solution of $\Gamma : y^2 z - x^3 - Axz^2 - Bz^3 = 0$. Otherwise, any point on E would correspond to multiple points on Γ. To rectify this degeneracy, projective spaces are used instead.

Definition A.2. *Let $n \in \mathbb{N}$ and $\mathbb{F}$ be an arbitrary field. Denote the n-fold Cartesian product $\mathbb{F} \times \mathbb{F} \times \cdots \times \mathbb{F}$ by $A_n(\mathbb{F})$. This is known as the **affine n-space** of $\mathbb{F}$.*

On any affine $n+1$-space where the origin has been removed, say $A_{n+1}(\mathbb{F}) \setminus \{(0, 0, \ldots, 0)\}$, one can introduce the relation $\sim$ defined by

$(x_1, x_2, \ldots, x_{n+1}) \sim (y_1, y_2, \ldots, y_{n+1})$ if and only if $(\lambda x_1, \lambda x_2, \ldots, \lambda x_{n+1}) = (y_1, y_2, \ldots, y_{n+1})$

for some non-zero $\lambda \in \mathbb{F} \setminus \{0\}$.

It is not difficult to show that this is an equivalence relation on $A_{n+1}(\mathbb{F}) \setminus \{(0, 0, \ldots, 0)\}$. Reflexivity follows from the fact $\mathbb{F}$ contains a multiplicative identity. Symmetry follows from the fact every non-zero element of $\mathbb{F}$ has an inverse. Finally, transitivity follows from the fact $\mathbb{F}$ is closed under multiplication. Therefore, $A_{n+1}(\mathbb{F}) \setminus \{(0, 0, \ldots, 0)\}$ can be partitioned into equivalence classes.

Definition A.3. *The **projective n-space** of $\mathbb{F}$, denoted $\mathbb{P}_n(\mathbb{F})$ is defined to be the set of all equivalence classes of $A_{n+1}(\mathbb{F})$ with respect to the equivalence relation $\sim$. The equivalence class of $(x_1, x_2, \ldots, x_{n+1})$ is denoted $[x_1, x_2, \ldots, x_{n+1}]$.*

As mentioned earlier, any solution to a homogeneous polynomial is independent of the choice of class representative i.e. if $(x_1, x_2, \ldots, x_{n+1})$ is a solution, then for any $\lambda \in \mathbb{F} \setminus \{0\}$, $(\lambda x_1, \lambda x_2, \ldots, \lambda x_{n+1})$ is a solution too. This implies that solutions to a homogeneous polynomial f are dependent only on the choice of equivalence class $[x_1, x_2, \ldots, x_{n+1}]$. In other words, the set

$$V_n(f) := \{[x_1, x_2, \ldots, x_{n+1}] \in \mathbb{P}_n(\mathbb{F}) \mid f(x_1, x_2, \ldots, x_{n+1}) = 0\}$$

is well-defined. The dependence of $V_n(f)$ on $\mathbb{F}$ is not made explicit.

Consider now the following lemma.

Lemma A.1. *Let H_n be the subset of $\mathbb{P}_n(\mathbb{F})$ in which every element satisfies $x_{n+1} = 0$ i.e.*

$$H_n := \{[x_1, x_2, \ldots, x_{n+1}] \in \mathbb{P}_n(\mathbb{F}) \mid x_{n+1} = 0\}.$$

Then there is a bijection between the affine n-space $A_n(\mathbb{F})$ and the projective complement $\mathbb{P}_n(\mathbb{F}) \setminus H_n$.

Proof. Define the map $\psi : \mathbb{P}_n(\mathbb{F}) \setminus H_n \longrightarrow A_n(\mathbb{F})$ by

$$\psi([x_1, x_2, \ldots, x_{n+1}]) = \left(\frac{x_1}{x_{n+1}}, \frac{x_2}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}\right).$$

Since $[x_1, x_2, \ldots, x_{n+1}] \notin H_n$, there is no issue in dividing by x_{n+1}.

Firstly, it is necessary to prove this map is well-defined. By definition, any other representation of the class $[x_1, x_2, \ldots, x_{n+1}]$ must be of the form $[\lambda x_1, \lambda x_2, \ldots, \lambda x_{n+1}]$ where $\lambda \neq 0$. Therefore,

$$\psi([\lambda x_1, \lambda x_2, \ldots, \lambda x_{n+1}]) = \left(\frac{\lambda x_1}{\lambda x_{n+1}}, \frac{\lambda x_2}{\lambda x_{n+1}}, \ldots, \frac{\lambda x_n}{\lambda x_{n+1}}\right) = \left(\frac{x_1}{x_{n+1}}, \frac{x_2}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}\right).$$

Thus, ψ is independent of the choice of representative and is well-defined.

Next, assume $\psi([x_1, x_2, \ldots, x_{n+1}]) = \psi([y_1, y_2, \ldots, y_{n+1}])$. This implies

$$\left(\frac{x_1}{x_{n+1}}, \frac{x_2}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}\right) = \left(\frac{y_1}{y_{n+1}}, \frac{y_2}{y_{n+1}}, \ldots, \frac{y_n}{y_{n+1}}\right).$$

Consequently, $x_i = \frac{x_{n+1}}{y_{n+1}} y_i$ for all $i \in \{1, 2, \ldots, n+1\}$ and it follows that

$$(x_1, x_2, \ldots, x_{n+1}) \sim (y_1, y_2, \ldots, y_{n+1})$$

and

$$[x_1, x_2, \ldots, x_{n+1}] = [y_1, y_2, \ldots, y_{n+1}].$$

Hence, ψ is an injective map.

Finally, to prove surjectivity, suppose $(x_1, x_2, \ldots, x_n) \in A_n(\mathbb{F})$. Then $[x_1, x_2, \ldots, x_n, 1] \in \mathbb{P}_n(\mathbb{F}) \setminus H_n$ and

$$\psi([x_1, x_2, \ldots, x_n, 1]) = (x_1, x_2, \ldots, x_n).$$

Therefore, ψ is surjective, and thus, bijective. □

This result leads to the following corollary.

Corollary. *Let*

$$V_2(\Gamma) := \{[a, b, c] \in \mathbb{P}_2(\mathbb{Q}) \mid \Gamma(a, b, c) = 0\}$$

and

$$V_2^*(\Gamma) := \{[a, b, c] \in \mathbb{P}_2(\mathbb{Q}) \mid \Gamma(a, b, c) = 0, c \neq 0\} = V_2(\Gamma) \setminus (V_2(\Gamma) \cap H_2).$$

Then, there exists a bijection between $V_2^*(\Gamma)$ and $E(\mathbb{Q})$.

Proof. Let $\psi^* : V_2^*(\Gamma) \longrightarrow A_n(\mathbb{Q})$ be the restriction of ψ to $V_2^*(\Gamma)$. It is immediate from the definition of Γ and the proof of lemma $A.1$ that ψ^* is well-defined, injective and the image of ψ^* is contained in $E(\mathbb{Q})$. Thus, it suffices to prove surjectivity. Suppose

$$\left(\frac{a}{c}, \frac{b}{c}\right) \in E(\mathbb{Q}).$$

Then, $\Gamma(a, b, c) = 0$ and $c \neq 0$. Hence $[a, b, c] \in V_2^*(\Gamma)$. However, by definition

$$\psi^*([a, b, c]) = \psi([a, b, c]) = \left(\frac{a}{c}, \frac{b}{c}\right).$$

Therefore, $E(\mathbb{Q}) \subset \mathrm{Im}(\psi^*)$, so ψ^* is a bijection between $V_2^*(\Gamma)$ and $E(\mathbb{Q})$. □

Therefore, $V_2^*(\Gamma)$ can be identified as a copy of $E(\mathbb{Q})$. Using this identification, the problem of extending $E(\mathbb{Q})$ to $E(\mathbb{Q}) \cup \{\infty\}$ can be reinterpreted as adding a unique element to $V_2^*(\Gamma)$. This is most naturally accomplished by extending $V_2^*(\Gamma)$ to $V_2(\Gamma)$ and introducing the identification

$$V_2(\Gamma) \longleftrightarrow E(\mathbb{Q}) \cup \{\infty\}.$$

Of course, it remains to show that $V_2^*(\Gamma)$ can be extended to $V_2(\Gamma)$ by adding a single element. To this end, consider the set

$$V_2(\Gamma)/V_2^*(\Gamma) = \{[a, b, c] \in \mathbb{P}_2(\mathbb{Q}) \mid \Gamma(a, b, c) = 0, c = 0\}.$$

Any element in this set must be of the form $[a, b, 0]$. Now, because $V_2(\Gamma) \subset \mathbb{P}_2(\mathbb{Q})$ and $[0, 0, 0] \notin \mathbb{P}_2(\mathbb{Q})$, a and b cannot both be zero. Assume then, without loss of generality, that $b \neq 0$. Then

$$(a, b, 0) = \left(\frac{a}{b}b, 1b, 0b\right).$$

Hence,

$$[a, b, 0] \sim \left[\frac{a}{b}, 1, 0\right].$$

However, since $\Gamma(a, b, 0) = 0$, this implies $a^3 = 0$, so $a = 0$ too. Therefore,

$$[a, b, 0] \sim \left[\frac{a}{b}, 1, 0\right] \sim [0, 1, 0]$$

and $V_2(\Gamma)/V_2^*(\Gamma)$ can contain at most one element i.e. $[0, 1, 0]$. It is easy to see that $V_2(\Gamma)/V_2^*(\Gamma)$ does indeed contain this element since

$$\Gamma(0, 1, 0) = 0 - 0^3 - 0^3 - 0^3 = 0.$$

It follows immediately that

$$V_2(\Gamma) = V_2^*(\Gamma) \cup V_2(\Gamma)/V_2^*(\Gamma) = V_2^*(\Gamma) \cup \{[0, 1, 0]\}.$$

Thus, ∞ merely represents the element $[0, 1, 0]$ added to the set $V_2^*(\Gamma)$ (interpreted as $E(\mathbb{Q})$).

Appendix B

P-adic Numbers

In this appendix, an account of p-adic numbers and a complete statement of Hensel's lemma are provided.

For an arbitrary prime p, the p-adic field $\mathbb{Q}_p$ is a completion of $\mathbb{Q}$ with respect to an absolute value depending only on p. The next few pages are dedicated to explaining these notions.

Firstly, it is necessary to discuss absolute values.

Definition B.1. *An **absolute value** of the field $\mathbb{F}$ is a map $|\cdot| : \mathbb{F} \longrightarrow \mathbb{R}_{\geq 0}$ satisfying the following conditions*

1. $|x| = 0 \Leftrightarrow x = 0$.
2. $|xy| = |x||y|$.
3. $|x + y| \leq |x| + |y|$.

An absolute value always induces a metric on the underlying field. If $|\cdot|$ is the absolute value in question, then

$$\mathrm{d}(x, y) := |x - y|$$

will be the induced metric.

Two distinct absolute values will induce equivalent metrics if they themselves are equivalent.

Definition B.2. *Two absolute values of the field* $\mathbb{F}$, $|\cdot|$ *and* $|\cdot|'$, *are said to be* ***equivalent*** *if there exists an exponent* $e > 0$ *such that*

$$|x|^e = |x|'$$

for all $x \in \mathbb{F}$.

Intuitively, a metric space (M, d) is complete if every sequence that could have a limit, *does* have a limit. In other words, any sequence that does not diverge or fluctuate erratically will converge to a point. Such sequences are called Cauchy sequences and are defined as follows.

Definition B.3. *A sequence* $\{x_i\}_{i=1}^{\infty}$ *in a metric space* (M, d) *is said to be* ***Cauchy*** *if for all* $\varepsilon > 0$, *there exists* $N \in \mathbb{N}$ *such that*

$$\mathrm{d}(x_i, x_j) < \varepsilon, \ \forall i, j > N.$$

With this notion in mind, we can now define a complete field.

Definition B.4. *Let* $|\cdot|$ *be an absolute value of the field* $\mathbb{F}$ *and* $\mathrm{d}(x, y) = |x - y|$ *be the corresponding metric. Then* $\mathbb{F}$ *is* ***complete*** *with respect to* $|\cdot|$ *if every Cauchy sequence in the metric space* $(\mathbb{F}, \mathrm{d})$ *converges to a limit in* $\mathbb{F}$.

Examples of complete fields are $\mathbb{R}$ and $\mathbb{C}$ which are complete with respect to the usual Euclidean norms. A non-example is the field $\mathbb{Q}$. Some rational Cauchy sequences do not converge to rational limits, e.g.

$$\frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \frac{13}{8}, \frac{21}{13} \cdots \longrightarrow \frac{1 + \sqrt{5}}{2}.$$

The incompleteness of $\mathbb{Q}$ makes it virtually impossible to do any sort of analysis on this field. To overcome this problem, it is common to work with completions/complete field extensions of $\mathbb{Q}$ instead.

Definition B.5. *Let $\mathbb{F}$ be a field with absolute value $|\cdot|$ and $\mathbb{K}$ be a complete field with absolute value $|\cdot|'$. Then $\mathbb{K}$ is the* ***completion*** *of $\mathbb{F}$ with respect to $|\cdot|$ if there exists an injective map*

$$\lambda : \mathbb{F} \longrightarrow \mathbb{K}$$

such that

1. *The absolute value is preserved i.e. $|\lambda(x)|' = |x|$ for all $x \in \mathbb{F}$.*
2. *The field $\mathbb{K}$ is the closure of $\lambda(\mathbb{F})$ with respect to the metric induced by $|\cdot|'$. In particular, $\lambda(\mathbb{F})$ is dense in $\mathbb{K}$.*

The prototypical example of a completion is $\mathbb{R}$, which is the completion of $\mathbb{Q}$ with respect to the usual Euclidean norm. For an example of how completions are constructed, see [**6**], pg 11.

In 1916, Alexander Ostrowski proved that, up to equivalence of absolute values, the only completions of $\mathbb{Q}$ are the real numbers $\mathbb{R}$ and the p-adic numbers $\mathbb{Q}_p$. The p-adic absolute value arises naturally from the study of prime divisors.

Definition B.6. *Let p be a prime and n be an integer. The* ***p-adic valuation*** *of n is defined to be*

$$\nu_p(n) := \begin{cases} \max\{k \in \mathbb{N} \cup \{0\} \mid p^k \text{ divides } n\} & \text{if } n \neq 0 \\ \infty & \text{if } n = 0. \end{cases}$$

The domain of this function can be extended from $\mathbb{Z}$ to $\mathbb{Q}$ via the functional relation

$$\nu_p\left(\frac{a}{b}\right) := \nu_p(a) - \nu_p(b).$$

It is easy to check this extended definition is well-defined.

The p-adic valuation satisfies the following properties for all $x, y \in \mathbb{Q}$:

1. $\nu_p(xy) = \nu_p(x) + \nu_p(y)$;
2. $\nu_p(x + y) \geq \min\{\nu_p(x), \nu_p(y)\}$.

Proof. Write $x = p^{k_1}\frac{a_1}{b_1}$ and $y = p^{k_2}\frac{a_2}{b_2}$ such that $k_1, k_2 \in \mathbb{Z}$ and $\gcd(a_i, p) = \gcd(b_i, p) = 1$ for $i \in \{1, 2\}$. This can always be accomplished by extracting prime powers from the numerators and denominators of x and y.

It follows immediately that $\nu_p(x) = k_1$ and $\nu_p(y) = k_2$. To prove the first property, multiply x and y together to get

$$xy = p^{k_1+k_2}\frac{a_1a_2}{b_1b_2}.$$

Since $\gcd(a_i, p) = \gcd(b_i, p) = 1$ for $i \in \{1, 2\}$, one must have $\gcd(a_1a_2, p) = \gcd(b_1b_2, p) = 1$. Therefore,

$$\nu_p(xy) = k_1 + k_2 = \nu_p(x) + \nu_p(y).$$

To prove the second property, assume first that $k_1 \leq k_2$ (by relabelling if necessary). Then

$$x + y = p^{k_1}\left(\frac{a_1}{b_1} + p^{k_2-k_1}\frac{a_2}{b_2}\right) = p^{k_1}\left(\frac{a_1b_2 + p^{k_2-k_1}a_2b_1}{b_1b_2}\right).$$

As before, we have $\gcd(b_1b_2, p) = 1$. Therefore,

$$\nu_p\left(\frac{a_1b_2 + p^{k_2-k_1}a_2b_1}{b_1b_2}\right) \geq 0$$

and consequently,

$$\nu_p(x + y) \geq k_1 = \min\{\nu_p(x), \nu_p(y)\}.$$

□

In order to construct the ***p*-adic absolute value** from the p-adic valuation, it suffices to exponentiate ν_p to the base p

$$\nu_p(x) \longrightarrow |x|_p := p^{-\nu_p(x)}.$$

It follows immediately from the definition of the p-adic valuation that $|x|_p \geq 0$ for all $x \in \mathbb{Q}$ and $|x|_p = 0$ if and only if $x = 0$. Furthermore, exponentiating the first property of the p-adic valuation guarantees that $|xy|_p = |x|_p|y|_p$. Therefore, in order to prove $|\cdot|_p$ is an absolute value on $\mathbb{Q}$, it suffices to prove

$$|x + y|_p \leq |x|_p + |y|_p.$$

Proof. Without loss of generality, assume $\nu_p(x) \leq \nu_p(y)$. Then

$$\nu_p(x + y) \geq \nu_p(x) = \min\{\nu_p(x), \nu_p(y)\}$$

by the second property of the p-adic valuation. Multiplying both sides by -1 gives

$$-\nu_p(x + y) \leq -\nu_p(x).$$

Next, exponentiate this inequality to get

$$p^{-\nu_p(x+y)} \leq p^{-\nu_p(x)}.$$

However,

$$p^{-\nu_p(x)} \leq \max\{p^{-\nu_p(x)}, p^{-\nu_p(y)}\}.$$

Therefore, one has

$$p^{-\nu_p(x+y)} \leq \max\{p^{-\nu_p(x)}, p^{-\nu_p(y)}\}$$

or equivalently,

$$|x+y|_p \leq \max\{|x|_p, |y|_p\}.$$

Since $|x|_p$ and $|y|_p$ are non-negative, we can apply the trivial inequality

$$\max\{|x|_p, |y|_p\} \leq |x|_p + |y|_p$$

to conclude that

$$|x+y|_p \leq |x|_p + |y|_p.$$

□

Thus, $|\cdot|_p$ is an absolute value of $\mathbb{Q}$. This absolute value is non-archimedean because it satisfies the ultra metric inequality.

We can now define the p-adic fields $\mathbb{Q}_p$.

Definition B.7. *The* ***p-adic field*** $\mathbb{Q}_p$ *is the completion of* $\mathbb{Q}$ *with respect to the p-adic absolute value* $|\cdot|_p$.

Using this definition, one can also construct p-adic integers. These are generalisations of $\mathbb{Z}$ contained in $\mathbb{Q}_p$. They are defined as follows.

Definition B.8. *The ring of p-**adic integers** $\mathbb{Z}_p$ is defined to be the subring of $\mathbb{Q}_p$ containing all elements with non-negative p-adic absolute values*

$$\mathbb{Z}_p := \{x \in \mathbb{Q}_p \mid |x|_p \leq 1\}.$$

To complete this section, three equivalent forms of Hensel's lemma are stated. A proof of Hensel's lemma may be found in [**6**], chapter 10.

The first version is the simplest to understand.

Hensel's Lemma. *Suppose $f(x) \in \mathbb{Z}_p[x]$ has a formal derivative $f'(x)$ and that there exists $\alpha \in \mathbb{Z}_p$ such that*

$$f(\alpha) \equiv 0 \mod (p),\ f'(\alpha) \not\equiv 0 \mod (p).$$

Then there exists a unique $\alpha' \in \mathbb{Z}_p$ satisfying

$$f(\alpha') = 0 \text{ and } \alpha' \equiv \alpha \mod (p).$$

Equivalently,

Hensel's Lemma. *Suppose $f(x) \in \mathbb{Z}_p[x]$ and that there exists $\alpha \in \mathbb{Z}_p$ such that*

$$|f(\alpha)|_p < 1,\ |f'(\alpha)|_p = 1.$$

Then there exists a unique $\alpha' \in \mathbb{Z}_p$ such that

$$f(\alpha') = 0 \text{ and } |\alpha' - \alpha|_p \leq |f(\alpha)|_p.$$

This version of the lemma is the one proved in [**6**].

Finally,

Hensel's Lemma. *Suppose $f(x) \in \mathbb{Z}_p[x]$ and that there exists $\alpha \in \mathbb{Z}_p$ such that*

$$|f(\alpha)|_p < |f'(\alpha)|_p^2.$$

Then there exists a unique $\alpha' \in \mathbb{Z}_p$ such that

$$f(\alpha') = 0 \text{ and } |\alpha' - \alpha|_p < |f'(\alpha)|_p.$$

Moreover,

$$|\alpha' - \alpha|_p = \left|\frac{f(\alpha)}{f'(\alpha)}\right|_p \text{ and } |f'(\alpha)|_p = |f'(\alpha')|_p.$$

The final version of Hensel's lemma is particularly important as it tells us exactly how far away the root α' is from the 'approximate root' α. The variable α' is sometimes referred to as the lift of α. Similar to Newton's approximation method, the proof of Hensel's lemma involves constructing better and better approximations for the roots of $f(x) \in \mathbb{Z}_p$.

Bibliography

[1] T. Apostol. *Modular Functions and Dirichlet Series in Number Theory*. Vol. 41 of *Graduate Texts in Mathematics.* 2nd ed., Springer-Verlag. (1990)

[2] M. Bhargava and A. Shankar. Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank 0. *Ann. Math.* 181, pg 587 - 621. (2015)

[3] C. Breuil, B. Conrad, F. Diamond and R. Taylor. On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises. *Jour. Amer. Math. Soc.* 14, pg 843 - 939. (2001)

[4] K. S. Brown. *Cohomology of Groups.* Vol. 87 of *Graduate Texts in Mathematics*. 2nd ed., Springer-Verlag. (1994)

[5] J. W. S. Cassels. Arithmetic on curves of genus 1. II. A general result. *J. Reine Angew. Math.* 203, pg 174 - 177. (1960)

[6] J. W. S. Cassels. *Lectures on Elliptic Curves*. Vol. 24 of *London Mathematically Society Student Texts*. Cambridge University Press. (1991)

[7] J. Coates and A. Wiles. On the conjecture of Birch and Swinnerton-Dyer. *Invent. Math.* 39, pg 223 - 251. (1977)

[8] P. Deligne. La conjecture de Weil I. *Publ. Math. IHÉS.* 43, pg 273 - 307. (1974)

[9] M. Deuring. Die Zetafunktion einer algebraischen Kurve von Geschlechte Eins I - IV. *Nachr. Akad. Wiss. Göttingen.*. (1953 - 1957)

[10] D. Goldfeld. Gauss' class number problem for imaginary quadratic fields. *Bull. Amer. Math. Soc.* 13, pg 23 - 37. (1985)

[11] R. Greenberg. On the Birch and Swinnerton-Dyer conjecture. *Invent. Math.* 72, pg 241 - 265. (1983)

[12] B. Gross and D. Zagier. Heegner points and derivatives of L-series. *Invent. Math.* 84, pg 225 - 320. (1986)

[13] R. Hartshorne. *Algebraic Geometry*. Vol. 52 of *Graduate Texts in Mathematics*. Springer-Verlag. (1977)

[14] H. Hasse. Zur Theorie der abstrakten elliptischen Funktionenkörper I, II, III. *J. Reine Angew. Math.* 175, pg 55 - 62, pg 69 - 88, pg 193 - 208. (1936)

[15] D. R. Heath-Brown. The average analytic rank of elliptic curves. *Duke Math. Jour.* 122, pg 591 - 623. (2004)

[16] G. Julia. Étude sur les formes binaires non quadratiques à indéterminées réelles, ou complexes, ou à indéterminées conjuguées. *Mem. Acad. Sci. l'Inst. France.* 55, pg 1 - 293. (1917)

[17] V. A. Kolyvagin. Finiteness of $E(\mathbb{Q})$ and $Ш_E$ for a subclass of Weil curves. *Izv. Akad. Nauk SSSR Ser. Math.* 52, pg 522 - 540. (1988)

[18] B. Mazur. Rational isogenies of prime degree. *Invent. Math.* 44, pg 129 - 162. (1978)

[19] L. J. Mordell. On the rational solutions of the indeterminate equations of the third and fourth degrees. *Proc. Camb. Phil. Soc.* 21, pg 179 - 19. (1922)

[20] K. Rubin. Tate-Shafarevich groups and L-functions of elliptic curves with complex multiplication. *Invent. Math.* 89, pg 527 - 560. (1987)

[21] G. Salmon. *Lessons Introductory to the Modern Higher Algebra*. 3rd ed., Hodges, Foster and Co. (1876)

[22] J. H. Silverman. *The Arithmetic of Elliptic Curves.* Vol. 106 of *Graduate Texts in Mathematics*. 2nd ed., Springer-Verlag. (1986)

[23] J. H. Silverman and J. Tate. *Rational points on Elliptic Curves*. Springer-Verlag. (1992)

[24] E. M. Stein and R. Shakarchi. *Princeton lectures in analysis II: Complex Analysis*. Princeton University Press. (2002)

[25] M. Stoll. Rational 6-cycles under iteration of quadratic polynomials. *Jour. London. Math. Soc. Comp. Math.* 11, pg 367 - 380. (2008)

[26] H. P. F. Swinnerton-Dyer and B. Birch. Notes on elliptic curves. I. *J. Reine Angew. Math.* 212, pg 7 - 25. (1963)

[27] H. P. F. Swinnerton-Dyer and B. Birch. Notes on elliptic curves. II. *J. Reine Angew. Math.* 218, pg 79 - 108. (1965)

[28] J. Tate. The arithmetic of elliptic curves. *Invent. Math.* 23, pg 179 - 206. (1974)

[29] J. B. Tunnell. A classical Diophantine problem and modular forms of weight $\frac{3}{2}$. *Invent. Math.* 72, pg 323 - 334. (1983)

[30] L. C. Washington. *Elliptic Curves: Number Theory and Cryptography.* 2nd ed., Chapman and Hall. (2008)

[31] A. Weil. On algebraic groups of transformations. *Amer. J. Math.* 77, pg 355 - 391. (1955)

[32] A. Weil. On algebraic groups and homogeneous spaces. *Amer. J. Math.* 77, pg 493 - 512. (1955)