The group structure on an elliptic curve

Nathan Alvarez Olson

Spring 2019

Bernd Siebert, Ph.D.
Department of Mathematics
Supervising professor

James Vick, Ph.D.
Department of Mathematics
Second reader

Department of Mathematics
Plan II Honors Program
Dean's Scholars Honors Program
The University of Texas at Austin

Abstract

An elliptic curve is defined most generally as the solution set $E(K)$ of a non-singular cubic polynomial f with coefficients in a field K. Via the theory of elliptic functions and the Weierstrass $\wp$-function specifically, a bijection is established between the complex torus and the $\mathbb{C}$-points of an elliptic curve. This map endows $E(\mathbb{C})$ with an abelian group structure, where the group law yields a nice geometric interpretation when the curve is embedded in $\mathbb{CP}^2$. Mordell's theorem, of which a special case is proved, implies $E(\mathbb{Q})$ is finitely generated. Lastly, the Nagell-Lutz theorem, which places a divisibility condition on the y-coordinates of points in G that implies G is finite, is proved.

Contents

1 Elliptic curves in projective space

Let $f \in K[x, y]$ be a polynomial in two variables over a field K. We say that the curve defined by $f(x, y) = 0$ is *non-singular* if the partial derivatives f_x and f_y never vanish simultaneously. Note that

$$\frac{dy}{dx} = \frac{\partial f/\partial x}{\partial f/\partial y}$$

Then, intuitively, a curve being non-singular means that a tangent line can always be found to a curve: we interpret the tangent line to be vertical when the denominator of the above expression vanishes [9, 26]. If this happens at the same time that the numerator vanishes, no such interpretation exists, and undesirable behavior (such as the curve having a self-intersection or a cusp) is exhibited.

Most generally, an *elliptic curve* is the solution set $E(K')$ (where K' could be an extension of the ground field K) to a non-singular cubic equation in two variables [13]. Normally, one will embed this solution set into projective space, as we will do shortly, which requires one to homogenize the cubic. Before doing this, some remarks on the different forms of an elliptic curve are in order. When working with elliptic curves, it is known that a general cubic equation can be transformed via birational transformations to an equation of the form

$$y^2 = f(x) = x^3 + ax^2 + bx + c$$

When $\operatorname{char} K \neq 2, 3$, we can do better, and write the curve as

$$y^2 = f(x) = x^3 + ax + b$$

[13]. Alternatively, when working with elliptic curves over $\mathbb{C}$, any elliptic curve can be written as

$$y^2 = 4x^3 - g_2x - g_3$$

which is sometimes referred to as *Weierstrass normal form* [6, p. 33]. Given the equivalence of these forms, we will often work in the form which most suits our needs, being sure to note which form we are using.

Then, if $y^2 = f(x)$, note that

$$\frac{dy}{dx} = \frac{-f'(x)}{2y}$$

by the aforementioned equation. In this form, note that nonsingularity is equivalent to f not having a double root, as otherwise, f and f' share that root. An equivalent definition and the one given in many sources, then, of an elliptic curve is the solution set to an equation of the form $y^2 = f(x)$, where $f(x)$ is in one of the aforementioned forms or is a general cubic and f and f' share no roots [6, p. 9] [12].

Before going further, there are two elliptic curves that bear mentioning as motivation. The first relates to the *congruent number problem*: a positive integer n is a *congruent number* if it is the area of a right triangle with rational sides. Though this may appear trivial at

first, note that 157 is a congruent number, yet the simplest right triangle has side lengths equal to

$$\frac{411340519227716149383203}{21666555693714761309610} \quad \text{and} \quad \frac{6803298487826435051217540}{411340519227716149383203}$$

[6, p.3–5]. As will be remarked upon at the end, the congruent number problem, to this day, is only partially resolved. This came as a consequence of difficult theorems concerning elliptic curves, as the congruent number problem can be formulated into the following statement regarding the elliptic curve C defined by $y^2 = x^3 - n^2x$: n is a congruent number if and only if the curve C has a rational point with x-coordinate the square of a rational number, an even denominator, and numerator relatively prime with n [1].

Another example of the power of elliptic curves is in the *Frey curve*, which appeared in the proof of Fermat's Last Theorem. Namely, for a solution $a^\varepsilon + b^\varepsilon = c^\varepsilon$ where $a, b, c \in \mathbb{Q}$ and ε is an integer greater than 2, one can define the elliptic curve $y^2 = x(x - a^\varepsilon)(x + b^\varepsilon)$. Ribet showed that such a curve, if it were to exist, would have peculiar properties that would contradict the *modularity theorem* (which we won't even try to define). Fermat's Last Theorem, then, was reduced to a special case of the modularity theorem, which was proved by Wiles in 1995 [14].

Elliptic curves over $\mathbb{C}$ (or, very similarly, $\mathbb{R}$) are often embedded in two-dimensional complex *projective space* $\mathbb{CP}^2$, which is defined as $(\mathbb{C}^3 - \{0\})/\sim$, where $\sim$ is the equivalence relation defined by $(x, y, z) \sim (x', y', z')$ if and only if there exists a nonzero $\lambda \in \mathbb{C}$ such that $(x, y, z) \sim (\lambda x', \lambda y', \lambda z')$ [11, p. 33]. The equivalence class of (x, y, z) is often written in *homogenous coordinates* as $[x : y : z]$. Note that $\mathbb{CP}^2$ can be thought of as $\mathbb{C}^2 \cup \mathbb{CP}^1$, as any point $[x : y : z]$ with $z \neq 0$ can be written as $[x/z : y/z : 1]$. Then, the set $\{[x : y : 0]\}$ is often called the "line at infinity" and the point $\mathcal{O} = [0 : 1 : 0]$—which will feature promptly in our coming discussion—is a distinguished point on this line.

To embed an elliptic curve $y^2 = x^3 + ax^2 + bx + c$ into $\mathbb{CP}^2$, one makes the variable substitution and multiplication

$$z^3\left[\left(\frac{y}{z}\right)^2 = \left(\frac{x}{z}\right)^3 + a\left(\frac{x}{z}\right)^2 + b\left(\frac{x}{z}\right) + c\right] \quad \Rightarrow \quad y^2z = x^3 + ax^2z + bxz^2 + cz^3$$

This is a specific example of *homogenizing* a polynomial. Note that the resulting polynomial equation is constant on the equivalence classes of $\mathbb{CP}^2$, and hence the solution set in $\mathbb{CP}^2$ is well-defined. Note that $\mathcal{O}$ is a solution to this equation, and all other solutions to this equation are of the form $[x : y : 1]$. As such, we will often denote the point $[x : y : 1]$ as (x, y), though one should not forget that these points lie in projective space.

[1] This equivalence follows from variable substitutions and algebraic manipulations of the equations $a^2 + b^2 = c^2$ and $ab/2 = n$ governing whether or not n is a congruent number. The details are a bit too complicated and a bit too far from the focus of this thesis to mention, though they can be found in [6, p.4-7] or on the Wikipedia page "Congruent number".

2 Elliptic functions

2.1 Elliptic functions and the Weierstrass $\wp$-function

Intricately related to elliptic curves are *elliptic functions*. As we will see, this relationship occurs via a specific elliptic function.

Fix $\omega_1, \omega_2 \in \mathbb{C}$ linearly independent over $\mathbb{R}$. A function $f\colon \mathbb{C} \to \mathbb{C}$ is *doubly periodic* with respect to ω_1 and ω_2 if $f(z + \omega_i) = f(z)$ for every $z \in \mathbb{C}$ and $i = 1, 2$. Another way of formulating this definition is in terms of lattices. Given $\omega_1, \omega_2 \in \mathbb{C}$, one can consider the *lattice* Λ in the complex plane consisting of the integral linear combinations of ω_1 and ω_2, i.e. $\omega_1\mathbb{Z} + \omega_2\mathbb{Z}$.

Note that for a doubly periodic function f, $f(z + \ell) = f(z)$ for every $\ell \in \Lambda$. This means that the behavior of such an f on the entire complex plane is completely determined by its behavior on its *fundamental parallelogram*, defined by $\Pi = \{c_1\omega_1 + c_2\omega_2 \mid 0 \leqslant c_1, c_2 \leqslant 1\}$, as any $z \in \mathbb{C}$ can be written as $\ell + p$, where $\ell \in \Lambda$ and $p \in \Pi$. This is analogous to singly periodic functions such as sin on $\mathbb{R}$.

A doubly periodic, meromorphic function is called an *elliptic function* relative to Λ [6, p. 14].

Proposition 2.1. *An elliptic function f which has no pole in Π must be constant.*

Proof. Because Π is compact, an elliptic function f without a pole in Π must be bounded on Π, and therefore all of $\mathbb{C}$. Because f is meromorphic, Liouville's theorem implies that f is constant [6, p. 15]. ■

Proposition 2.2. *For $\lambda \in \mathbb{C}$, consider the translation $\Pi' = t + \Pi$ of the period parallelogram. If an elliptic function f has no poles on the boundary $\partial\Pi'$ of Π', then the sum of residues of $f(z)$ in Π' is zero.*

Proof. The residue theorem implies that the sum of the residues of f within Π' is given by

$$\frac{1}{2\pi i}\int_{\partial\Pi'} f\ dz$$

But this integral is zero, as f takes identical values on opposite sides of Π' by the doubly periodic condition, and the path of integration is in opposite directions on opposite sides of Π' [6, p. 15]. ■

Because f is meromorphic, it has finitely many poles in any bounded region. Then, one may always choose some t for which $t + \Pi$ has no poles on the boundary [6, p. 15]. This observation allows us to state

Proposition 2.3. *Suppose that f has no zeros or poles on the boundary of $\Pi' = t + \Pi$. If $\{m_i\}$ and $\{n_i\}$ are the orders of the zeros and poles in Π', respectively, then $\sum m_i = \sum n_i$.*

Proof. Consider the "logarithmic derivative" f'/f. Let p be a zero or pole. Then,

$$f(z) = a_m(z-p)^m + \dots$$
$$f'(z) = a_m m(z-p)^{m-1} + \dots$$

where m is the order of the zero or pole (negative if a pole). Note then that

$$f'(z)/f(z) = m(z-a)^{-1} + \dots$$

Hence, f'/f has a pole where f has a zero or pole, this pole is simple, and the residue is equal to the order of the zero or pole (negative if it is a pole). By the previous proposition, summing the residues gives $\sum_i m_i + \sum_i -n_i = 0$ [6, p. 16]. ■

Here is a lemma which will prove useful later:

Lemma 2.4. *For f an elliptic function, $\Pi' = t + \Pi$ a translation of the fundamental period of Π of Λ with no zeros or poles on the boundary, and $\{a_i\}$ and $\{b_j\}$ the set of zeros and poles in Π' each counted as many times as its multiplicity, $\sum a_i - \sum b_j \equiv 0 \pmod{L}$.*

Proof. Let s be a zero or pole with order m ($m < 0$ if s is a pole). As in the proof of the previous proposition, the logarithmic derivative f'/f has an expansion starting with $m(z-s)^{-1}$ around s. Then the expansion of $zf'(z)/f(z)$ around s starts with

$$z\left(m(z-s)^{-1} + \dots\right) = (s + (z-s))\left(m(z-s)^{-1} + \dots\right) = sm(z-s)^{-1} + \dots$$

The sum of the residues inside Π', then, is $\sum_s m_s s = \sum a_i - \sum b_j$. The residue theorem implies this is equal to

$$\frac{1}{2\pi i}\int_{\partial\Pi'} \frac{zf'(z)}{f(z)}\ dz$$

We calculate this path integral over the opposite sides of Π' [2]:

$$\frac{1}{2\pi i}\left(\int_t^{t+\omega_2} \frac{zf'(z)}{f(z)}\ dz - \int_{t+\omega_1}^{t+\omega_1+\omega_2} \frac{zf'(z)}{f(z)}\ dz\right) = \frac{-\omega_1}{2\pi i}\int_t^{t+\omega_2} \frac{f'(z)}{f(z)}\ dz$$

after using the u-substitution $z \mapsto z + \omega_1$ and subtracting the integrands. Making the u-substitution $u = f(z)$ and letting C be the path traced by f as z ranges $t \to t + \omega_2$, this integral simplifies to

$$\frac{-\omega_1}{2\pi i}\int_C \frac{1}{u}\ du$$

But this integral is simply the $-\omega_1$ times the winding number of C around 0, which is an integer n_1. Similarly, the integral over the other pair of opposite sides of Π'—i.e. $t \to t + \omega_1$ and $t + \omega_2 \to t + \omega_1 + \omega_2$—is equal to $-n_2\omega_2$ for some integer n_2. Hence, $\sum a_i - \sum b_j = -n_1\omega_1 - n_2\omega_2 \equiv 0 \pmod{L}$ [6, p. 30–1]. ■

[2] Noting that $\Pi = \{c_1\omega_1 + c_2\omega_2 \mid 0 \leqslant c_1, c_2 \leqslant 1\}$.

We now define our most important example of an elliptic function.

Definition 2.1. [6, p. 16]. For Λ a lattice, the *Weierstrass $\wp$-function*, denoted $\wp(z;\Lambda)$, or by $\wp(z)$ when the lattice is clear from context, is given by

$$\wp(z;\Lambda) = \frac{1}{z^2} + \sum_{\substack{\ell\in\Lambda \\ \ell\neq 0}} \frac{1}{(z-\ell)^2} - \frac{1}{\ell^2}$$

Proposition 2.5. *The sum defining $\wp(z)$ converges absolutely and uniformly for z in any compact subset of $\mathbb{C}-\Lambda$.*

Showing this requires from the following lemmas:

Lemma 2.6.

$$\sum_{\substack{\ell\in\Lambda \\ \ell\neq 0}} \|\ell\|^{-s} < \infty$$

for $s > 2$.

Proof. We first claim that the number of points of Λ in the annulus defined by $n-1 \leqslant |\ell| \leqslant n$ is of order magnitude n.

Let $D > 0$ be a number larger than the distance between any two points in Π, and let I_n be the collection of fundamental parallelograms which have nontrivial intersection with the anulus [3]. Note that each parallelogram in I_n is contained in the larger anulus A_n defined by $n-1-D \leqslant |\ell| \leqslant n+D$. This follows from the choice of D. Then, if $i_n = \|I_n\|$, we have that

$$i_n * \text{Area}(\Pi) \leqslant \text{Area}(A_n) \leqslant \pi\left[(n+D)^2 - (n-1-D)^2\right] = \pi\left[2n(1+2D) - 1 - 2D\right] \leqslant Cn$$

for some C. Hence, $i_n \leqslant C'n$ for some C', as required.

We are now equipped to prove the lemma. Write the sum over Λ as a sum over the points of Λ in each annulus A_n for $n = 1, 2, \ldots$:

$$\sum_{\ell\neq 0} \|\ell\|^{-s} = \sum_{n=1}^{\infty} \sum_{n-1\leqslant\|\ell\|<n} \|\ell\|^{-s} \leqslant C' \sum_{n=1}^{\infty} nn^{-s}$$

which converges for $s > 2$ [12] [6, p. 17]. ■

Proof of proposition. The summand of $\wp$ can be rewritten as $(2z\ell - z^2)/((z-\ell)^2\ell^2)$.

We show absolute and uniform convergence by comparison with the series $|\ell|^{-3}$. Using our assumption on the compact set being away from the points of Λ, let m be the maximum value of $|z|$ on our compact set. Then, we consider the sum for $|\ell| \geqslant 2m$, where we are ignoring finitely many terms. Using the triangle inequality and the reverse triangle inequality,

$$|2\ell - z| \leqslant |2\ell| + |z| \leqslant |2\ell| + \frac{5|\ell|}{2} = \frac{5|\ell|}{2}$$

[3] By "fundamental parallelograms", I mean translations of Π by elements of Λ

$$|z-\ell| \geqslant ||z|-|\ell|| \geqslant \left|\frac{|\ell|}{2}-|\ell|\right| = \frac{|\ell|}{2}$$

Then, for $|\ell| \geqslant 2m$, we have

$$\left|\frac{1}{(z-\ell)^2}-\frac{1}{\ell^2}\right| = \left|\frac{2z\ell-z^2}{(z-\ell)^2\ell^2}\right| = \frac{|z||2\ell-z|}{|z-\ell|^2|\ell|^2} \leqslant \frac{m\frac{5|\ell|}{2}}{\frac{|\ell|^2}{2^2}|\ell|^2} = \frac{10m}{|\ell|^3}$$

Absolute and uniform convergence follow from comparison with the series $|\ell|^{-3}$ [12] [6, p.17].

■

Proposition 2.7. *$\wp$ is an elliptic function. Its only pole is a double pole at each point of Λ.*

Proof. Fix $\ell \in \Lambda$. Then, the function $\wp(z)-(z-\ell)^{-2}$ is continuous at $z=\ell$ by the same logic as the previous proposition. The absolute and uniform convergence of the series defining $\wp$, which consists of holomorphic functions, therefore implies $\wp$ is differentiable away from the points of Λ. These two facts imply $\wp$ is meromorphic with a double pole at each point of Λ.

To show that $\wp$ is an elliptic function, first note $\wp$ is even. Using a reindexing $\ell \mapsto -\ell$ of Λ yields

$$\wp(-z) = \frac{1}{(-z)^2}+\sum_{\substack{\ell\in\Lambda\\ \ell\neq 0}}\frac{1}{(-z-\ell)^2}-\frac{1}{\ell^2} = \frac{1}{z^2}+\sum_{\substack{-\ell\in\Lambda\\ \ell\neq 0}}\frac{1}{(z-\ell)^2}-\frac{1}{\ell^2} = \wp(z)$$

Second, note that $\wp'$ can be found by differentiating term-wise:

$$\wp' = -2\sum_{\ell\in\Lambda}(z-\ell)^3$$

$\wp'$ is doubly periodic (and an elliptic function, as it is meromorphic), as $\wp'(z+\ell)=\wp'(z)$ for any $\ell \in \Lambda$, simply by reindexing the sum.

To show that $\wp$ is an elliptic function, it suffices to show that $\wp(z+\omega_i)=\wp(z)$ for each $i=1,2$. Consider first ω_1. Then, the derivative of $\wp(z+\omega_1)-\wp(z)$ is $\wp'(z+\omega_1)-\wp(z)$, which is 0 by the double periodicity of $\wp'$. Hence, $\wp(z+\omega_1)-\wp(z)=C$ for some constant C. Letting $z=-\omega_1/2$, we have $C=\wp(-\omega_1/2)-\wp(-\omega_1/2)=0$ by the evenness of $\wp$. The argument for ω_2 is identical. The conclusion follows [6, p.17–8].

■

2.2 The roles of $\wp$ and $\wp'$

Let's investigate the behavior of $\wp$ and $\wp'$. Fix $u \in \mathbb{C}$. Because $\wp-u$ has a double pole at 0, proposition 2.3 implies that $\wp-u$ has either two simple zeros or a double zero in a translation Π' of Π, and therefore in Π.

Consider η, a "half-lattice point", i.e.

$$\eta \in \left\{\frac{\omega_1}{2}, \frac{\omega_2}{2}, \frac{\omega_1+\omega_2}{2}\right\}$$

Then, $\wp'(\eta - z) = \wp(-\eta - z) = -\wp'(\eta + z)$ by the double periodicity and oddness of $\wp'$. Continuity of $\wp'$ means η is a zero of $\wp'$. These zeros are all simple and are the only zeros of $\wp'$ in Π by proposition 2.3: $\wp'$ has its only pole in Π at 0 and this pole is of order 3. As such, $\wp - u$ has two simple zeros for each u, except for when u is $\wp(\eta)$. This exception occurs because $\wp - u$ has a zero at η, and $(\wp - \wp(\eta))'(\eta) = \wp'(\eta) = 0$ implies this is a double zero of $\wp - \wp(\eta)$.

Note that the collection $\{\wp(\eta)\}$ is distinct: if $\wp(\eta_1) = \wp(\eta_2)$, then $\wp - \wp(\eta_1)$ has double zeros at η_1 and η_2, contradicting the fact that it only has one double pole.

In summary, $\wp$ obtains every value of $\mathbb{C} \cup \{\infty\}$ twice on $\mathbb{C}/\Lambda$, except for the four points $\{\wp(\eta) | \eta \in H\} \cup \{\infty\}$. These points have only one preimage in $\mathbb{C}/\Lambda$ [6, p.21].

This fact will be used in our proof of

Proposition 2.8. *$\wp$ and $\wp'$ generate the field of elliptic functions for Λ.*

Similar to how Fourier's theorem shows periodic functions can be expressed as weighted sums of sine and cosine functions, this proposition implies doubly periodic functions on Λ can be expressed as a rational expression in $\wp$ and $\wp'$.

The key to the proof of the proposition is the following lemma:

Lemma 2.9. *The subfield of even elliptic functions for a lattice Λ is generated by $\wp$.*

Proof. Let f be an even elliptic function. For this proof only, let Π be the fundamental parallelogram with two sides removed: $\Pi = \{\lambda_1\omega_1 + \lambda_2\omega_2 | 0 \leqslant \lambda_i < 1\}$. Note that every element of $\mathbb{C}$ can be written uniquely as $\ell + p$, where $\ell \in \Lambda$ and $p \in \Pi$.

Let a be a zero of f of order m in Π where a is not a half-lattice point and $a \neq 0$. Define a^* to be the point "symmetric" to a: $a^* = \omega_1 + \omega_2 - a$ if a is in the interior of Π and $a^* = \omega_1 - a$ or $a^* = \omega_2 - a$ if a is on the side of ω_1 or ω_2, respectively.

Because f is doubly periodic and even, $f(a^* - z) = f(-a - z) = f(a + z)$. Also, a is a zero of order m, so $f(a + z) = a_m z^m + \dots$. Hence, a^* is also a zero of order m, as

$$f(a^* + z) = f(a - z) = a_m(-z)^m + \dots$$

Suppose now that a is a half-lattice point. As before, $f(a - z) = f(a + z)$, and so

$$a_m(-z)^m + \dots = f(a - z) = f(a + z) = a_m z^m + \dots$$

It follows that m is even.

If b is a pole of f of order m, the same logic implies that b^* is a pole of the same order if b is not a half-lattice point, and that b has even order otherwise.

Let $\{a_i\}$ be the zeros of f in Π, listed as follows:

- From each pair of symmetrical zeros a and a^*, choose one and list it as many times as its multiplicity.
- List each a that is a half-lattice point half as many times as its multiplicity.

Let $\{b_j\}$ be the same list, except for the poles of f. Note firstly that 0 does not appear in either of these lists, as we have ignored it for the time being. Also note $\wp(a_i)$ and $\wp(b_j)$ are defined for each i and j as all of the a_i's and b_j's are nonzero. Define

$$g(z) = \frac{\prod_i(\wp(z) - \wp(a_i))}{\prod_j(\wp(z) - \wp(b_j)}$$

Note that g is a rational function of $\wp$. If we show that g and f have the same zeros and poles—with multiplicities—as f, then proposition 2.1 implies that $f = cg$ for some constant c and the proposition is proved.

We do this first for the nonzero zeros and poles. Because 0 is the only pole of $\wp$, the nonzero zeros and poles of g must come from the zeros of its numerator and denominator, respectively. Specifically, the zeros of g correspond to the zeros of $\wp(z) - \wp(a_i)$ and the poles to the zeros of $\wp - \wp(b_j)$.

Our earlier characterization of the behavior of $\wp$ means $\wp(z) - \wp(a)$ has a zero at a. If a is a half-lattice point, this is a double zero; otherwise, it has two simple zeros, one at a and the other at a^* by the evenness and double periodicity of $\wp$. Because of how we created the lists $\{a_i\}$ and $\{b_i\}$, we find that the orders of these zeros match the orders of the zeros and poles of f.

It remains to show that order of the zero or pole at 0 of g matches that of f. This follows simply from proposition 2.3: for an elliptical function with $\{n_i\}$ and $\{m_j\}$ the orders of the poles and zeros respectively, we have that $\sum n_i - \sum m_j = 0$. This holds for both f and g. But the quantities $\sum n_i - \sum m_j$ for f and g are identical except for one term: the order of the zero or pole (negative if a pole) at 0. But since these expressions are equal, the order of the zero or pole at 0 must be the same in f and g.

This concludes the proof [6, p.19–20]. ■

The proposition now follows easily:

Proof of proposition. The functions

$$f_1 = \frac{f(z) + f(-z)}{2} \quad \text{and} \quad f_2 = \frac{f(z) - f(-z)}{2\wp'(z)}$$

are even and therefore rational functions in $\wp$ by the lemma. Observing that $f = f_1 + \wp' f_2$ completes the proof [6, p.18].

■

Corollary 2.10. *$(\wp')^2$ is a cubic polynomial in $\wp$.*

Proof. $\wp'$ is odd with a triple pole at 0 and three simple zeros at the three half-lattice points, as discussed earlier. Hence, $(\wp')^2$ is even with a pole of order 6 at 0 and three double zeros. This means there are three a_i's and zero b_j's. The corollary follows [6, p.20]. ■

By the previous corollary, we write $(\wp')^2 = f(\wp)$, where we know from the proof of lemma 2.9 that

$$f(x) \propto \prod_{\eta \text{ a half-lattice point}} (\wp(z) - \wp(\eta))$$

and hence f has distinct roots. We are interested in an expression for f of the form $ax^3 + bx^2 + cx + d$, and now endeavor to find expressions (dependent on Λ) for a, b, c and d. These can be found by expanding the Laurent expressions for $\wp$, $\wp'$, and their powers around $z = 0$. Suppose we do so in a disc of radius rc, where $r < 1$ and c is the minimum of $|\ell|$ for nonzero $\ell \in \Lambda$.

Fixing a nonzero $\ell \in \Lambda$ and noting that $(1-x)^{-2} = 1 + 2x + 3x^2 + \ldots$, the summand in the expression for $\wp$ can be written as

$$\frac{1}{(z-\ell)^2} + \frac{1}{\ell^2} = \frac{1}{\ell^2}\left(\frac{1}{(1-z/\ell)^2} - 1\right) = 2\frac{z}{\ell^3} + 3\frac{z^2}{\ell^4} + \ldots$$

Hence,

$$\wp(z) = z^{-2} + \sum_{\substack{\ell \in \Lambda \\ \ell \neq 0}} 2\frac{z}{\ell^3} + 3\frac{z^2}{\ell^4} + \cdots + (i-1)z^{i-2}\ell^{-i} + \ldots$$

This sum is absolutely convergent in the aforementioned disc: note first

$$\left\|2\frac{z}{\ell^3}\right\| + \left\|3\frac{z^2}{\ell^4}\right\| + \cdots = 2\|z\|\|\ell\|^{-3}\left(1 + \frac{3}{2}\left\|\frac{z}{\ell}\right\| + \frac{4}{2}\left\|\frac{z}{\ell}\right\|^2 + \ldots\right)$$

Because $\|z\| < rc < r\|\ell\|$ for every $\ell \in \Lambda$, this is strictly less than

$$2\|z\|\|\ell\|^{-3}\left(1 + \frac{3}{2}r + \frac{4}{2}r^2 + \ldots\right) < \frac{2\|z\|}{(1-r)^2}\|\ell\|^{-3}$$

which converges because $r < 1$. Absolute convergence follows by comparison with $\sum |\ell|^{-3}$. Absolute convergence validates the following change in the order of summation:

$$\wp(z) = \frac{1}{z^2} + \sum_{\substack{\ell \in \Lambda \\ \ell \neq 0}} \sum_{i=3}^{\infty} (i-1)z^{i-2}\ell^{-i} = \frac{1}{z^2}\sum_{i=3}^{\infty}(i-1)z^{i-2}G_i$$

where

$$G_i := \sum_{\substack{\ell \in \Lambda \\ \ell \neq 0}} \ell^{-i}$$

is a well-defined constant for each i that depends solely on Λ. Also note that $G_i = 0$ when i is odd, as $(-\ell)^{-i} = -\ell^{-i}$. We can use this modified formula for $\wp$ to write expansions:

$$\begin{aligned}
\wp(z) &= z^{-2} + 3G_4 z^2 + 5G_6 z^4 + \dots \\
(\wp(z))^2 &= z^{-4} + 6G_4 + 10G_6 z^2 + \dots\dots \\
(\wp(z))^3 &= z^{-6} + 9G_4 z^{-2} + 15G_6 + \dots \\
\wp'(z) &= -2z^{-3} + 6G_4 z + 20G_6 z^3 + \dots \\
(\wp'(z))^2 &= 4z^{-6} - 24G_4 z^{-2} - 80G_6 + \dots
\end{aligned}$$

Writing $(\wp')^2 = a\wp^3 + b\wp^2 + c\wp + d$ and equating coefficients of the z^{-6}, z^{-4}, z^{-2}, and z^0 terms gives $a = 4$, $b = 0$, $c = -60G_4$, $d = -140G_6$. As such, we have that $(\wp')^2 = f(\wp)$, where—defining $g_2 := 60G_4$ and $g_3 := 140G_6$ as is tradition—$f = 4x^3 - g_2 x - g_3$ [6, p.22–4].

This cubic relationship (which is in Weierstrass normal form!) and the fact that $f \in \mathbb{C}[x]$ allows us to define a bijective correspondence between our complex lattice $\mathbb{C}/\Lambda$ and the elliptic curve defined by $y^2 = f(x)$ in Weierstrass normal form embedded in $\mathbb{CP}^2$. It is given explicitly by

$$\begin{aligned}
\varphi\colon \mathbb{C}/\Lambda &\to \mathbb{CP}^2 \\
0 &\mapsto \mathcal{O} \\
z &\mapsto [\wp(z) : \wp'(z) : 1]
\end{aligned}$$

Note that $\varphi(\mathbb{C}/\Lambda)$ is contained in the elliptic curve by the fact that $(\wp')^2 = f(\wp)$.

φ is surjective. Every point of the form $[x : 0 : 1]$ in $E(\mathbb{C})$ is mapped to by one of the three half lattice points in $\mathbb{C}/\Lambda$. Note that the roots of f are distinct and these three half lattice points are in bijective correspondence with the roots of f. Any point of the form $[x : \pm y : 1]$ with $y \neq 0$ is mapped to by two distinct values $\pm z \pmod{\Lambda}$, as $\wp'(-z \bmod \Lambda) = -\wp'(z)$.

Lastly, $\mathcal{O}$ is mapped to by $(0, 0) \in \mathbb{C}/\Lambda$ only.

On the other hand, the injectivity of φ follows from the fact $\wp$ is a two-to-one map to $\mathbb{C} \cup \{\infty\}$ except for the four values previously mentioned [6, p.24].

2.3 The group law on an elliptic curve

Being in bijective correspondence with the abelian group $\mathbb{C}/\Lambda$ endows $E(\mathbb{C})$ with an abelian group structure. Though we will not show it, any nonsingular complex elliptic curve can be transformed into Weierstrass normal form $y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$, where Λ is some lattice [9, p.43].

This group structure will be the focus of the rest of our discussion.

Quite interestingly, the group attached to an elliptic curve by φ possesses a meaningful geometric interpretation. Speaking imprecisely, to add two points $P_1 + P_2$, one first draws a line through the two points P_1 and P_2, tangent if $P_1 = P_2$. Then, $P_1 + P_2$ is equal to the third point of intersection of this line with the elliptic curve (counting multiplicities),

reflected across the x-axis [4].

Let's make this rigorous. Note first that $\varphi(0) = \mathcal{O}$ is the additive identity, and so we know how to add $P_1 + P_2$ when either P_1 or P_2 is $\mathcal{O}$, though thinking of lines intersecting with the point at infinity can be hard to imagine. Furthermore, $-(x, y) = (x, -y)$, as

$$\varphi(-z) = [\wp(-z) : \wp'(-z) : 1] = [\wp(z) : -\wp'(z) : 1] = (x, -y)$$

Geometrically, the third point of intersection of a vertical line through two points (or one point if the line is a tangent) with the same x-coordinate is the "point at infinity" $\mathcal{O}$.

We have shown that this geometric description holds for $\mathcal{O}$ or when $P_1 = -P_2$. Otherwise, when $P_1, P_2 \neq \mathcal{O}$ and $P_1 \neq -P_2$, the line L connecting P_1 and P_2 is of the form $y = mx + b$.

A point $P = (\wp(z), \wp'(z))$ is on L if and only if it is a zero of $\wp'(z) - (m\wp(z) + b)$. Note that $\wp' - m\wp - b$ has three zeros in $\mathbb{C}/\Lambda$, as it has a triple pole at 0. Similarly, by Bézout's theorem, we know that L has three intersections (counting multiplicities) with the elliptic curve. We now show that the multiplicity of a zero z_i of $\wp' - m\wp - b$ agrees with the multiplicity of the intersection of L with the elliptic curve at $P_i = \varphi(z_i)$.

Let $z_1, z_2, -z_3$ be the three zeros, included as many times as their multiplicity, of $\wp' - m\wp - b$. None of these is the negative of another one as L is not a vertical line. Note that $-z_1, -z_2, z_3$ are the zeros of $\wp' + m\wp + b$ by the evenness and oddness of $\wp$ and $\wp'$, respectively. Then, the six zeros of

$$(\wp' - m\wp - b)(\wp + m\wp + b) = (\wp')^2 - (\wp + b)^2$$

are $\{\pm z_i\}$. Then,

$$(\wp')^2 - (\wp + b)^2 = f(\wp) - (\wp + b)^2 = 4\prod_{i=1}^{3}(\wp - x_i)$$

where the x_i are the roots of $f(x) - (mx + b)$. Note that the multiplicity of x_i is the multiplicity of the intersection of L with the elliptic curve at the point $(x_i, f(x_i))$. Assume without loss of generality that $\wp(\pm z_1) = x_1$, as $\wp$ maps $\{\pm z_i\} \to \{x_i\}$ as sets by the fact that the z_i's are zeros. The multiplicity of x_1 is the number of x_2, x_3 which equal x_1, which is equal to the number of $\pm z_2, \pm z_3$ which are equal to $\pm z_1$. But this is the number of $z_2, -z_3$ which equal z_1, i.e. the multiplicity of z_1. Hence, the multiplicity of a zero of $\wp' - m\wp - b$ is the multiplicity of the intersection of L with the curve.

We now show that $-(P_1 + P_2) = \varphi(-(z_1 + z_2))$ lies on L, which happens iff $-(z_1 + z_2)$ is a zero of $\wp' - m\wp - b$. Two of its zeros are z_1 and z_2, and the third is $-(z_1 + z_2) \pmod{\Lambda}$ by lemma 2.4. To use the lemma, we need the above fact about the multiplicities of the z_i's, as we did not know *a priori* that z_1 and z_2 were of the same order as the multiplicity of the intersection of P_1 and P_2 with the curve, which by assumption added to 3. Now that they together have multiplicity two (if $z_1 = z_2$, then it is of at least multiplicity two), we can use the lemma.

[4] Note that any line will intersect the elliptic curve three times, counting multiplicities, as a consequence of *Bezout's theorem*. In our case, this follows from the more simple fact that $\mathbb{C}$ is algebraically closed, so $4x^3 - g_2x - g_3 - (mx + b)^2$, where $y = mx + b$ is the line connecting through one or more of the points on the curve, has three roots.

To conclude, $\varphi(-(z_1+z_2)) = -\varphi(z_1+z_2)$, so the third point of intersection of L with the curve is $-(P_1+P_2)$. As such, our geometric description of the addition law on $E(\mathbb{C})$ is correct [6, p.32–3].

We will have use for explicit formulas for the addition law. Note that we have only endowed elliptic curves of the form $y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$, or curves which can be linearly transformed to this form, with this structure. However, as discussed, this is not concerning.

As such, let $y^2 = ax^3 + bx^2 + cx + d$ be a general elliptic curve over K, a field with odd characteristic, as the formulas will work in this general of a setting. Let $P_1, P_2 \in E(K) - \mathcal{O}$ with $P_1 \neq -P_2$. Then, the tangent line is of the form $y = \lambda x + \mu$, where

$$\lambda = \begin{cases} \frac{y_2-y_1}{x_2-x_1} & \text{if } P_1 \neq P_2 \\ \frac{dy}{dx} = \frac{f'(x_1)}{2y_1} & \text{else} \end{cases}$$

Note that dy/dx can be found by implicitly differentiating $y^2 = f(x)$, and $f'(x)$ is the formal derivative in $K[x]$. Then, x_1, x_2, and x_3 are the three roots of $f(x) - (\lambda x + \mu)^2$. Note that

$$f(x) - (\lambda x + \mu)^2 = ax^3 + (b - \lambda^2)x^2 + \dots$$

The sum of the roots is equal to the negative of the coefficient of the second highest term divided by the leading coefficient. Hence, $x_1 + x_2 + x_3 = -(b - \lambda^2)/a$. It follows that

$$\begin{aligned} x_3 &= -x_1 - x_2 - \frac{b}{a} + \frac{1}{a}\lambda^2 \\ y_3 &= -y_1 + \lambda(x_1 - x_3) \end{aligned}$$

Note that one could define this addition structure on an elliptic curve over K and verify mechanically that the group axioms hold. By using the elliptic functions $\leftrightarrow$ elliptic curves correspondence, we are able to avoid this tedious procedure.

We will also have use for the "duplication formula" for $2(x, y)$ for $f(x) = x^3 + ax^2 + bx + c$ later, so we derive it now:

$$\begin{aligned} x(2(x,y)) &= -2x - a + \left(\frac{3x^2 + 2ax + b}{2y}\right)^2 \\ &= \frac{4y^2(-2x - a) + (3x^2 + 2ax + b)^2}{4y^2} \\ &= \frac{x^4 - 2bx^2 - 8cx + b^2 - 4ac}{4y^2} \end{aligned}$$

3 The Mordell theorem

Our main goal in this section is to prove the celebrated *Mordell theorem*, which states that for an elliptic curve defined by rational coefficients, $E(\mathbb{Q})$ is finitely generated. This will be a

long process, and unfortunately we will not be able to prove the statement in full generality. The bulk of the work in proving the Mordell theorem—as we will see—is in proving what is often called the "weak Mordell theorem", which states the index $[E(\mathbb{Q}) : 2E(\mathbb{Q})]$ of $2E(\mathbb{Q})$ in $E(\mathbb{Q})$ is finite. If one believes this fact, then we can prove the Mordell theorem in full generality.

3.1 Height on an elliptic curve

The proof of the Mordell theorem, assuming the weak Mordell theorem, follows by introducing a *height function* on $E(\mathbb{Q})$, which we do so now.

Given a rational number p/q written in lowest terms, define $H\colon \mathbb{Q} \to [0, \infty)$ given by $p/q \mapsto \max\{|p|, |q|\}$. Overloading the letter H, define

$$\begin{aligned} H\colon E(\mathbb{Q}) &\to [0, \infty) \\ (x, y) &\mapsto H(x) \\ \mathcal{O} &\mapsto 1 \end{aligned}$$

Which H is intended will be clear from context. The height function will ultimately be used in proving the Mordell theorem will be $h = \log \circ H$.

We now state and prove the following theorem, from which the Mordell theorem will follow. After that, it remains only to show that h has the required properties, and that $[E(\mathbb{Q})\colon 2E(\mathbb{Q})] < \infty$.

Descent theorem 3.1. *Let G be an abelian group where $[G : 2G] < \infty$. If $h\colon G \to [0, \infty)$ is a function satisfying*

1. *For any real number M, there exist only finitely many $g \in G$ satisfying $h(g) \leqslant M$.*
2. *For every $g_0 \in G$, there exists a constant κ_0 so that $h(g + g_0) \leqslant 2h(g) + \kappa_0$ holds for every $g \in G$.*
3. *There exists a constant κ so that $h(2g) \geqslant 4h(g) - \kappa$ holds for every $g \in G$.*

then G is finitely generated.

Proof. Let $Q_1, \ldots, Q_n$ be a complete list of representatives of the cosets of $2G$ in G. By our second condition on h, there exists a constant κ_i for each $i = 1, \ldots, n$ such that $h(g - Q_i) \leqslant 2h(g) + \kappa_i$ holds for every $g \in G$. Let $\kappa' = \max\{\kappa_i\}$. Then $h(g - Q_i) \leqslant 2h(g) + \kappa'$ holds for every $i = 1, \ldots, n$ and $g \in G$.

Fix $g \in G$. It is in some coset, say the i_1-th coset. We write $g - Q_{i_1} = 2g_1$ for some $g_1 \in G$. We do the same with g_1, yielding a series of statements

$$\begin{aligned} g_1 - Q_{i_2} &= 2g_2 \\ g_2 - Q_{i_3} &= 2g_3 \\ &\vdots \\ g_{m-1} - Q_{i_m} &= 2g_m \end{aligned}$$

Let κ denote the constant from our third condition. This condition gives, for each $j = 2, \ldots, m$,

$$4h(g_j) \leqslant h(2g_j) + \kappa = h(g_{j-1} - Q_{i_j}) + \kappa \leqslant 2h(g_{j-1}) + \kappa' + \kappa$$

This can be rewritten as

$$h(g_j) \leqslant \frac{3}{4}h(g_{j-1}) - \frac{1}{4}\left(h(g_{j-1}) - \kappa' - \kappa\right)$$

If $h(g_{j-1}) \geqslant \kappa' + \kappa$, then this becomes $h(g_j) \leqslant 3h(P_{j-1})/4$.

This is a long-winded way of saying that if we start with an element $g \in G$ of height larger than some constant (specifically, $\kappa' + \kappa$), then our sequence $g_1, g_2, \ldots$ decreases in height faster than $(3/4)^i$.

Consider the sequence of statements for g_i. The first equation implies $g = Q_{i_1} + 2g_1$. The second gives an expression for g_1, and the third one for g_2, and so on. Substituting these in gives

$$g = Q_{i_1} + 2Q_{i_2} + \cdots + 2^{m-1}Q_{i_m} + 2^m g_m$$

By the fact that $h(g_i) \to 0$ as $i \to \infty$, there exists a finite m such that $h(g_m) \leqslant \kappa + \kappa'$. But by the first condition, the set $S = \{g \in G | h(g) \leqslant \kappa' + \kappa\}$ is finite. As such, our expression for g tells us that $\{Q_i\} \cup S$, a finite set, generates G [1].

■

It remains to show that h has the claimed properties.

Lemma 3.2. *For any real number M, there exist only finitely many g satisfying $h(g) \leqslant M$.*

Proof. Write $p/q \in \mathbb{Q}$ in reduced terms. There are only finitely many p and q satisfying $|p|, |q| \leqslant e^M$. The lemma follows. ■

Lemma 3.3. *For every $g_0 \in E(\mathbb{Q})$, there exists a constant κ_0 so that $h(g + g_0) \leqslant 2h(g) + \kappa_0$ holds for every $g \in E(\mathbb{Q})$.*

Proof. Let $(x, y) = (m/M, n/N)$ be a rational point written in lowest terms on the elliptic curve $y^2 = x^3 + ax^2 + bx + c$. We first show that denominators of x and y are related.

Substituting the values of x and $'y$ into the governing equation yields

$$M^3 n^2 = N^2 m^3 + aN^2 M m^2 + bN^2 M^2 m + cN^2 M^3$$

We see that $N^2|M^3n^2$, but because N and n are relatively prime, $N^2|M^3$ in $\mathbb{Z}$. The equation above also implies

$$M^3n^2 - aN^2Mm^2 - bN^2M^2m - cN^2M^3 = N^2m^3$$

Using the relatively prime condition on M and m shows that $M|N^2$. Substituting $N^2 \propto M$ into the equation shows that $M^2|N^2m^3$ and hence $M|N$. Substituting $N \propto M$ gives that $M^3|N^2m^3$, implying $M^3|N^2$.

We conclude $M^3 = N^2$. After defining $e = N/M$, we may write that $x = m/e^2$ and $y = n/e^3$. Note as well that $m \leqslant H(x,y)$ and $n \leqslant \sqrt{H(x,y)}$. Substituting the values of x and y into our curve equation and clearing the denominators gives

$$n^2 = m^3 + am^2e^2 + bme^4 + ce^6$$

Therefore,

$$n^2 \leqslant |m^3| + |am^2e^2| + |bme^4| + |ce^6| \leqslant H(x,y)^3\left(1 + |a| + |b| + |c|\right)$$

implying that $n \leqslant \mu H(x,y)^{3/2}$ for some μ.

We are now able to prove the lemma. If we prove that the inequality holds for any (x_0, y_0) outside of a finite set of points, then it holds for all points, as we can always take κ_0 to be the maximum of the heights of the ignored points and the κ_0 we established on all but these finitely many points.

Without loss of generality, assume (x_0, y_0) is not equal to $\pm(x,y)$ or $\mathcal{O}$ and let (α, β) be the sum of these two points. Under our assumption on (x_0, y_0), the explicit formula for the addition law given previously yields

$$\alpha = \left(\frac{y - y_0}{x - x_0}\right)^2 - a - x_0 - x = \frac{Ay + Bx^2 + Cx + D}{Ex^2 + Fx + G}$$

for some constants $A, B, \ldots, G$ that may depend on x_0 and y_0. Note that neither x^3 nor y^2 occur in the numerator, as they eliminate each other when the expression is simplified. Substituting our expressions for x and y, clearing the denominators, taking the height, and using the triangle inequality yields

$$H(\alpha,\beta) = H(\alpha) \leqslant \max\left\{|Ane| + |Bm^2| + |Cme^2| + |De^4|, |Em^2| + |Fme^2| + |Ge^4|\right\}$$

Our bounds on m, n, and e give a further upper bound:

$$H(x,y)^2 \max\left\{|A\mu| + |B| + |C| + |D|, |E| + |F| + |G|\right\}$$

Applying log gives $h(g + g_0) \leqslant 2h(x,y) + \kappa_0$ [3] [9, p.68–71]. ■

Lemma 3.4. *There exists a constant κ so that $h(2g) \geqslant 4h(g) - \kappa$ holds for every $g \in E(\mathbb{Q})$.*

Proof. As with the previous proof, we can safely ignore finitely many cases. Without loss of generality, then, we assume $2g \neq \mathcal{O}$ and write $2(x, y) = (\alpha, \beta)$.

From the duplication formula given earlier

$$\alpha = \frac{x^4 - 2bx^2 - 8cx + b^2 - 4ac}{4x^3 + 4ax^2 + 4bx + 4c}$$

Note that the numerator and denominator do not vanish simultaneously in $\mathbb{C}$ by the nonsingularity of the elliptic curve. Hence, the proof is concluded by this next lemma: ■

Lemma. *Suppose $f, g \in \mathbb{Z}[x]$ do not not vanish simultaneously in $\mathbb{C}$, $d = \max(\deg f, \deg g)$, and $r \in \mathbb{Q}$ such that $g(r) \neq 0$. Then, $h\left(\frac{f(r)}{g(r)}\right) \geqslant dh(r) - \kappa$ for some κ dependent on f and g.*

Proof. Choose $u, v \in \mathbb{Q}[x]$ such that $uv + vg = 1$. If $r = m/n$ in lowest terms, write $F = n^d f$ and $G = n^d g$. Note that $F(r), G(r) \in \mathbb{Z}$ and $u(r)F(r) + v(r)G(r) = n^d$.

Let $e = \max\{\deg u, \deg v\}$ and define A to be the least common multiple of the denominators of the coefficients of u and v and. Then, $An^e u(r), An^e v(r) \in \mathbb{Z}$. Hence,

$$An^e u(r)F(r) + An^e u(r)G(r) = An^{d+e}$$

By Bézout's identity, $\gcd(F(r), G(r))$ divides An^{d+e}.

Suppose without loss of generality that $f = a_0 x^d + \cdots + a_d$ is the polynomial of degree d. $F(r) = a_0 m^d + a_1 n m^{d-1} + \cdots + a_d n^d$ gives

$$F(r) + n\left(-a_1 m^{d-1} - \cdots - a_d n^{d-1}\right) = a_0 m^d$$

Because n and m are relatively prime, Bézout's identity once again implies $\gcd(n, F(r))$ divides a_0. Looking then at the common factors of n and $F(r)$ and using $\gcd(F(r), G(r)) | An^{d+e}$, we see that $\gcd(F(r), G(r))$ divides Aa_0^{d+e}.

Then,

$$H\left(\frac{f(r)}{g(r)}\right) = H\left(\frac{F(r)}{G(r)}\right) = \frac{1}{\gcd(F(r), G(r))} \max\{|F(r)|, |G(r)|\}$$

is bounded below by $\max\{|F(r)|, |G(r)|\} / Aa_0^{d+e}$. As such,

$$\frac{H\left(\frac{f(r)}{g(r)}\right)}{H(r)^d} \geqslant \frac{\max(|F(r)|, |G(r)|)}{Aa_0^{d+e} \max\{|m|^d, |n|^d\}} = \frac{\max(|f(r)|, |g(r)|)}{\frac{1}{|n|^d} Aa_0^{d+e} \max\{|m|^d, |n|^d\}} = \frac{\max(|f(r)|, |g(r)|)}{Aa_0^{d+e} \max\{|r|^d, 1\}}$$

As $r \to \infty$, the RHS approaches some finite, nonzero limit. This quantity is also always positive, as f and g do not vanish simultaneously by our assumption. As such, the RHS is bounded below by a positive constant $1/C$. It follows that

$$h\left(\frac{f(r)}{g(r)}\right) = \log H\left(\frac{f(r)}{g(r)}\right) \geqslant \log CH(r)^d = dh(r) - \log(1/C)$$

[3] ■

3.2 The weak Mordell theorem: the maps φ and ψ

All that remains in proving the Mordell theorem is to prove $[E(\mathbb{Q}) : 2E(\mathbb{Q})]$ is finite. This, however, will take more effort than the previous lemmas combined. Furthermore, though it holds in general, we lack the machinery to prove it, and will assume that $f(x)$ has a rational root (or, equivalently, a rational point of order 2) which we will denote T.

With this assumption, we change variables and move the rational point of order 2 to the origin, making $f(x)$ of the form $x^3 + ax^2 + bx$.

The following lemma serves as the motivation for our proof of the weak Mordell theorem, and will ultimately be the last step in the proof.

Lemma 3.5. *For abelian groups A and B equipped with homomorphisms $\varphi\colon A \to B$ and $\psi\colon B \to A$ satisfying $\psi \circ \varphi(a) = 2a$, $\varphi \circ \psi(b) = 2b$ and $|A\colon \psi(B)|, |B\colon \varphi(A)| < \infty$, the index $|A\colon 2A|$ satisfies $|A\colon 2A| \leqslant |A\colon \psi(B)||B\colon \varphi(A)|$.*

Proof. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$ be representatives of $A/\psi(B)$ and $B/\varphi(A)$, respectively. Let $a \in A$. Then, there exist representatives a_i and b_j and elements $b \in B$ and $a' \in a$ such that $a - a_i = \psi(b)$ and $b - b_j = \varphi(a')$. Then,

$$a = a_i + \psi(b) = a_i + \psi(b_j + \varphi(a')) = a_i + \psi(b_j) + 2a'$$

and hence the finite set $\{a_i + \psi(b_j)\}$ contains a full set of representatives of $A/2A$. The conclusion follows [9, p.87–8]. ∎

The group A will be the group of rational points $E(\mathbb{Q})$ for our curve C, which we will denote by G in this section, while B, as we will see, will be the rational points on a related curve $\overline{C}$. We now construct these maps φ and ψ.

If C is the plane curve defined by $y^2 = x^3 + ax^2 + bx$, define $\overline{C}$ by $y^2 = x^3 + \overline{a}x^2 + \overline{b}x$, where $\overline{a} = -2a$ and $\overline{b} = a^2 - 4b$. Note that $\overline{\overline{C}}$ is defined by $y^2 = x^3 + 4ax^2 + 16bx$, and $(x, y) \in \overline{\overline{C}}$ if and only if $(x/4, y/8) \in C$. Loosely, $\overline{\overline{C}}$ and C are "essentially the same" and the map $\overline{}$ serves as a "dualizing" map; more precisely, they're isomorphic under the map $(x, y) \mapsto (x/4, y/8)$ [5].

Define

$$\begin{aligned} \varphi\colon C &\to \overline{C} \\ \{T, \mathcal{O}\} &\mapsto \overline{\mathcal{O}} \\ (x, y) &\mapsto \left(\frac{y^2}{x^2}, y\frac{x^2 - b}{x^2}\right) \end{aligned}$$

Note that the codomain of φ is as claimed: if $x \neq 0$, then

$$\varphi(x)^3 - 2a\varphi(x)^2 + (a^2 - 4b)\varphi(x) = \frac{y^2}{x^2}\left(\frac{y^4}{x^4} - 2a\frac{y^2}{x^2} + a^2 - 4b\right)$$

[5] The fact that this is a group homomorphism is easily checked with the addition formulas.

simplifies to

$$\frac{y^2}{x^4}\left(\frac{(y-ax^2)^2-4bx^4}{x^2}\right)=\frac{y^2}{x^4}\left(\frac{(x^3+bx)^2-4bx^4}{x^2}\right)=\left(\frac{y(x^2-b)}{x^2}\right)^2=\varphi(y)^2$$

[9, p.77]. φ is a group homomorphism. Let $P_1=(x_1,y_1), P_2=(x_2,y_2)\in G$ be nonzero and that they're not both $(0,0)$, as otherwise $\varphi(P_1+P_2)=\varphi(P_1)+\varphi(P_2)$ follows trivially. Suppose first that $P_2=(0,0)$. As $\varphi(P_2)=\overline{\mathcal{O}}$, we seek to prove $\varphi(P_1+P_2)=\varphi(P_1)$. The formula for the group law gives

$$x_3=-x_1-a+\left(\frac{-y_1}{-x_1}\right)^2=\frac{b}{x_1}\quad,\quad y_3=-y_1+\frac{-y_1}{-x_1}\left(x_1-\frac{b}{x_1}\right)=\frac{-by_1}{x_1^2}$$

Then,

$$\varphi(x_3,y_3)=\left(\left(\frac{-by_1/x_1}{b/x_1}\right)^2,\frac{\left(\frac{-by_1}{x_1^2}\right)\left(\frac{b^2}{x_1^2}-b\right)}{\left(\frac{b}{x}\right)^2}\right)=\left(\frac{y_1^2}{x_1^2},y_1\frac{x_1^2-b}{x_1^2}\right)=\varphi(x_1,x_1)$$

Furthermore, for any $P\in G$, $\varphi(-P)=-\varphi(P)$ follows directly from the formulas.

Now, let $P_1+P_2+P_3=\mathcal{O}$ for $P_1,P_2,P_3\notin\{\mathcal{O},T\}$. Note that it suffices to show $\varphi(P_1)+\varphi(P_2)+\varphi(P_3)=\overline{\mathcal{O}}$, as then

$$\varphi(P_1+P_2)=\varphi(-P_3)=-\varphi(P_3)=\varphi(P_1)+\varphi(P_2)$$

Denote $\varphi(P_i)=(\overline{x}_i,\overline{y}_i)$ and let $\lambda x+\mu$ be the line through P_1,P_2 and P_3. $P_i\neq T$ for each i implies $\mu\neq 0$. To show that $\sum\varphi(P_i)=\overline{\mathcal{O}}$ is to show that they all lie on some line $y=\overline{\lambda}x+\overline{\mu}$. Note that this only suffices if the $\varphi(P_i)$'s are distinct: normally, we must show that the $\overline{x}_i$'s are the roots of $(\overline{\lambda}x+\overline{\mu})^2-\overline{f}(x)$ in order to account for intersection multiplicities, as we did when first confirming the geometric definition of the group law on $E(\mathbb{C})$. However, φ is a continuous map. Once it is established as a homomorphism when the $\varphi(P_i)$'s are distinct, continuity means it is a homomorphism for all the points of C.

Using that $\mu\neq 0$, we define

$$\overline{\lambda}=\frac{\mu\lambda-b}{\mu}\quad,\quad\overline{\mu}=\frac{\mu^2-a\mu\lambda+b\lambda^2}{\mu}$$

Then, P_i is on the line defined by $y=\overline{\lambda}x+\overline{\mu}$, as

$$\begin{aligned}\overline{\lambda}\overline{x}_i+\overline{\mu}&=\frac{(\mu\lambda-b)y_i^2+(\mu^2-a\mu\lambda+b\lambda^2)\,x_i^2}{\mu x_i^2}\\&=\frac{\mu\lambda(y_i^2-ax_i^2)-b(y_i-\lambda x_i)(y_i+\lambda x_i)+\mu^2x_i^2}{\mu x_i^2}\\&=\frac{x_i^2(\lambda x_i+\mu)-by_i}{x_i^2}=\frac{y_i(x_i^2-b)}{x_i^2}=\overline{y}_i\end{aligned}$$

We conclude that φ is a homomorphism [9, p.80–1]. One can apply the $\overline{\ }$ map to get a map $\overline{\varphi}\colon \overline{C} \to \overline{\overline{C}}$. Let ψ denote the composition of $\overline{\varphi}$ with the isomorphism $\overline{\overline{C}} \leftrightarrow C$ given by $(x, y) \mapsto (x/4, y/8)$. Graphically, our maps are as follows:

$$C \xrightarrow{\varphi} \overline{C} \xrightarrow{\overline{\varphi}} \overline{\overline{C}} \overset{(x/4,y/8)}{\longleftrightarrow} C \xrightarrow{\varphi} \overline{C}, \qquad \overline{C} \xrightarrow{\psi} C$$

Note that $\overline{\varphi}$ is a homomorphism because it is defined identically to φ, but with $\overline{a}$ and $\overline{b}$ taking the place of a and b, and so is ψ.

It remains to show that $\psi \circ \varphi$ is multiplication by two. We abbreviate this process, as the math is exceptionally tedious. First note that the x-coordinate in the duplication formula can be written as $(x^2-b)^2/4y^2$. Skipping the calculations, the y-coordinate in the duplication formula is given by

$$y(2(x, y)) = \frac{(x^2 - b)\,(x^4 + 2ax^3 + 6bx^2 + 2abx + b^2)}{8y^3}$$

On the other hand,

$$\psi \circ \varphi(x, y) = \psi\left(\frac{y^2}{x^2}, \frac{y(x^2 - b)}{x^2}\right) = \left(\frac{(x^2 - b)^2}{4y^2}, \frac{(x^2 - b)\,(y^4 - (a^2 - 4b)x^4)}{8x^2y^3}\right)$$

simplifies to

$$\left(\frac{(x^2 - b)^2}{4y^2}, \frac{(x^2 - b)\,(x^4 + 2ax^3 + 6bx^2 + 2abx + b^2)}{8y^3}\right)$$

As such, $\psi \circ \varphi(x, y) = 2(x, y)$. Similarly, note that $\varphi \circ \psi(\varphi(P)) = \varphi(2P) = 2\varphi(P)$. But $\varphi\colon C \to \overline{C}$ is surjective, so in fact $\varphi \circ \psi(\overline{x}, \overline{y}) = 2(\overline{x}, \overline{y})$ for all $(\overline{x}, \overline{y}) \in \overline{C}$.

Note that we have only show that these two compositions are a doubling map for points expressible as (x, y) with $x \neq 0$. However, φ and ψ both map $\mathcal{O}$ and T to $\mathcal{O}$ (or $\overline{\mathcal{O}}$), so in fact they are doubling maps for all points of $E(\mathbb{Q})$ [9, p.82].

3.3 Finite index and the map α

It remains to show that the indices of $\varphi(G)$ in $\overline{G}$ and $\psi(\overline{G})$ in G are finite. To do this, we will need a description of $\varphi(G)$ [9, p.83–5]:

- Clearly, $\overline{\mathcal{O}} \in \varphi(G)$.
- $(0, 0) \in \varphi(G)$ iff $\overline{b}$ is a square. Using the equation for $\overline{C}$, notice that $(0, 0) \in \varphi(G)$ iff there exists a rational point $(x, y) \in C$ with $x \neq 0$ and $y = 0$. But $y = 0$ iff $x(x^2 + ax + b) = 0$ iff $\overline{b} = a^2 - 4b$ is a square, by the quadratic formula.
- When $\overline{x} \neq 0$, $(\overline{x}, \overline{y}) \in \varphi(G)$ iff $\overline{x}$ is a square in $\mathbb{Q}$. The forward direction is obvious. Now suppose that $\overline{x} = r^2$. For $i = 1, 2$, define

$$(x_i, y_i) = \left(\frac{1}{2}\left(r^2 - a + (-1)^{i-1}\frac{\overline{y}}{r}\right), (-1)^{i-1}x_i r\right)$$

Note

$$x_1 x_2 = \frac{1}{4}\left((r^2-a)^2 - \frac{\overline{y}^2}{r^2}\right) = \frac{1}{4}\left(\frac{\overline{x}^3 - 2a\overline{x}^2 + a^2\overline{x} - y^2}{\overline{x}}\right) = b$$

We claim that $(x_i, y_i) \in C$ and $\varphi(x_i, y_i) = (\overline{x}, \overline{y})$.

Showing that $(x_i, y_i) \in C$ is equivalent to showing that $y_i^2/x_i^2 = x_i + a + b/x_i$. Using the definition of y_i and that $x_1 x_2 = b$, this is equivalent to showing that $r^2 = x_1 + a + x_2$, which follows easily from the definition of x_i. Lastly, note that $\varphi(x_i, y_i) = (\overline{x}, \overline{y})$, as $y_i^2/x_i^2 = \overline{x}$ by definition, and

$$\frac{y_i(x_i^2 - b)}{x_i^2} = \frac{(-1)^{i-1}x_i r(x_i^2 - x_1 x_2)}{x_i^2} = r(x_1 - x_2) = \overline{y}$$

We now prove that $[G : \psi(\overline{G})] < \infty$. We omit the proof that $[\overline{G} : \varphi(G)] < \infty$, as it proceeds similarly.

Proposition 3.6. *Let $\mathbb{Q}^*$ denote the multiplicative group of $\mathbb{Q}$. Define*

$$\begin{aligned} \alpha \colon G &\to \mathbb{Q}^*/\mathbb{Q}^{*2} \\ (x, y) &\mapsto x \pmod{\mathbb{Q}^{*2}} \\ \mathcal{O} &\mapsto 1 \pmod{\mathbb{Q}^{*2}} \\ T &\mapsto b \pmod{\mathbb{Q}^{*2}} \end{aligned}$$

Then,

1. *α is a group homomorphism.*
2. *If $p_1, \ldots, p_k$ are the distinct primes dividing b, the image of α is contained in the subgroup $\{\pm p_1^{\varepsilon_1} \ldots p_i^{\varepsilon_i} | \varepsilon_i = 0, 1\}$.*
3. *$\ker \alpha = \psi(\overline{G})$.*

From the last two statements, α induces an injective map from $G/\psi(\overline{G})$ into a finite subgroup. As such, the proof of this proposition finishes the proof of the weak Mordell theorem.

Proof. (Part 1) Note that $\alpha(-(x, y)) = \alpha(x, -y) = x$ which is the inverse of x in $\mathbb{Q}^*/\mathbb{Q}^{*2}$. As before, it suffices to show that when $P_1 + P_2 + P_3 = \mathcal{O}$ for $P_i \in E(\mathbb{Q})$, then $\prod \alpha(P_i) = 1$.

Firstly, the cases when any of the $P_i = \mathcal{O}$ are trivial.

The case that all $P_i = T$ is impossible, as T is of order 2. The case that two $P_i = T$ is trivial: $T + T + P_3 = \mathcal{O}$ implies $P_3 = \mathcal{O}$. Then, without loss of generality, suppose $P_1 = T$. Because the line through the P_i's passes through $(0, 0)$, $\mu = 0$. Factoring out x, we get x_2 and x_3 are the zeros of $\lambda^2 x = x^2 + ax + b$. The fact that the product of the roots of a polynomial is equal to the constant term yields

$$\alpha(P_1)\alpha(P_2)\alpha(P_3) = bx_1x_2 = b^2 \equiv 1 \pmod{\mathbb{Q}^{*2}}$$

Now, let $P_1 + P_2 + P_3 = \mathcal{O}$ where $P_i \neq \mathcal{O}, T$. Then, they lie on some line $y = \lambda x + \mu$. The x-coordinates of these points are the roots of the equation $x^3 + ax^2 + bx - (\lambda x + \mu)^2$. The product of the roots is

$$\alpha(P_1)\alpha(P_2)\alpha(P_3) = x_1x_2x_3 = \mu^2 \equiv 1 \pmod{\mathbb{Q}^{*2}}$$

(Part 2). By the proof of lemma 3.3, we write a rational point as $(m/e^2, n/e^3)$ where $\gcd(m, e) = 1$. Because $\alpha(m/e^2, n/e^3) = m \pmod{\mathbb{Q}^{*2}}$, the image of α is precisely the possible residues of m mod $\mathbb{Q}^{*2}$. This is generated by the prime factors of m which occur to an odd power.

As in lemma 3.3, substituting the expressions for x and y into $y^2 = f(x)$ and clearing denominators yields

$$n^2 = m^3 + am^2e^2 + bme^4 = m(m^2 + ame^2 + be^4)$$

Each prime divisor of m that occurs to an odd power in m must also occur to an odd power in $m^2 + ame^2 + be^4$. But then it must occur to an odd power in be^4 and therefore b by the fact that m and e are relatively prime.

(Part 3). Earlier we described $\varphi(G)$. Note however that the same description applies to $\overline{\varphi}(\overline{G})$ as long as $\overline{b}$ is replaced with $\overline{\overline{b}}$, and hence to ψ, as x is a square in $\mathbb{Q}$ iff $x/4$ is. But the description given for $\psi(\overline{G})$ is exactly the description of $\ker \alpha$ [3] [9, p.85–7]. ■

4 Points of finite order

The classification of finitely generated abelian groups and Mordell's theorem tells us that $E(\mathbb{Q}) \simeq \mathbb{Z}^r \oplus \mathrm{Tor}(E(\mathbb{Q}))$, where r is the *rank* of $E(\mathbb{Q})$ and $\mathrm{Tor}(E(\mathbb{Q}))$ is the *torsion subgroup* of points of finite order in $E(\mathbb{Q})$.

What can we say about $\mathrm{Tor}(E(\mathbb{Q}))$? When we consider the larger field $\mathbb{C}$, the answer is easy: the points of order dividing n in $\mathbb{C}/\Lambda$ is precisely

$$\left\{ \frac{a}{n}\omega_1 + \frac{b}{n}\omega_2 \middle| a, b = 0, 1, \ldots, n-1 \right\}$$

Hence, via the isomorphism $\mathbb{C}/\Lambda \simeq E(\mathbb{C})$, $\mathrm{Tor}(E(\mathbb{C})) \simeq (\mathbb{Z}/n\mathbb{Z})^2$.

$E(\mathbb{Q})$ is much smaller, and the rational points of finite order can be found by a finite process, as guaranteed by

The Nagell-Lutz theorem 4.1. *Suppose $y^2 = f(x) = x^3 + ax^2 + bx + c$ is a nonsingular elliptic curve where $f \in \mathbb{Z}[x]$. Let $(x, y) \in \mathrm{Tor}(E(\mathbb{Q}))$. Then, $x, y \in \mathbb{Z}$, and if $y \neq 0$, y^2 divides the discriminant D of f, given by*

$$D = -4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2$$

Proving this theorem requires new definitions and the introduction of some notation. Fix a prime p. Then, every element of $\mathbb{Q}$ can be written as $p^n a/b$, where a, b, and p are relatively prime. Let the *p-adic valuation*, denoted by ν_p or simply ν when p is clear from context, be the map $\mathbb{Q}^* \to \mathbb{Z}$ given by $p^n a/b \mapsto n$.

As per usual, write a rational point (x, y) on $y^2 = x^3 + ax^2 + bx + c$ as $(x/e^2, y/e^3)$. Suppose this can be written as

$$\left(\frac{n_x}{d_x p^{\varepsilon_x}}, \frac{n_y}{d_y p^{\varepsilon_y}} \right)$$

where $\varepsilon_x, \varepsilon_y > 0$. Then, $\varepsilon = 2\varepsilon$ and $\varepsilon_y = 3\varepsilon$ for some $\varepsilon \geqslant 0$. This motivates the following definition of a collection of subsets of $E(\mathbb{Q})$ for each p and indexed by $\mathbb{Z}^{>0}$:

$$C(p^\varepsilon) := \{(x, y) \in E(\mathbb{Q}) \mid \nu_p(x) \leqslant -2\varepsilon \text{ and } \nu_p(y) \leqslant -3\varepsilon\} \cup \{\mathcal{O}\}$$

Note that these subsets form a descending chain:

$$E(\mathbb{Q}) \supset C(p) \supset C(p^2) \supset \dots$$

Proving the first part of the Nagell-Lutz theorem—that $\mathrm{Tor}(E(\mathbb{Q})) \subset \mathbb{Z}^2$—amounts to showing that $(x, y) \notin C(p)$ for any prime p.

One can think of these $C(p^n)$'s as smaller and smaller neighborhoods of $\mathcal{O}$ in the p-adic topology. This serves as motivation for defining the change of coordinates $t = x/y$ and $s = 1/y$ [6]. Then, $x = t/s$ and $y = 1/s$, and our equation $y^2 = x^3 + ax^2 + bx + c$ becomes $s = t^3 + at^2 s + bts^2 + cs^3$ after clearing denominators. The reason for such a coordinate change is that it maps the "point at infinity" $\mathcal{O}$ to the origin and points (x, y) with large x- and y-values close to the origin. If $\mathcal{O} = [0 : 1 : 0] = [0 : N : 0]$ in homogeneous coordinates, where N can be arbitrarily large, then

$$t = \frac{x}{y} = 0 \quad , \quad s = \frac{1}{y} = \frac{1}{N} \to 0$$

shows the reasoning for $\mathcal{O} \mapsto (0, 0)$.

On the other hand, points with $y = 0$ are not mapped anywhere. This is not concerning, however, as points with $y = 0$ are of finite order, and have integer coordinates by the fact that $f(x) \in \mathbb{Z}[x]$ is monic, and $\mathbb{Z}$ is integrally closed in $\mathbb{Q}$. From here on, we can safely ignore points of order 2, though we will be careful to note when we are doing so.

To summarize, we have a map $(x, y) \mapsto (t, s)$ which is a bijection between $E(\mathbb{Q})$ minus the points of order 2 and the points in the (t, s)-plane.

[6] What happens to $\mathcal{O}$ and points with $y = 0$ will be addressed shortly.

Furthermore, we can mimic the group law in the (x, y)-plane in the (t, s)-plane, for if $y = \lambda x + \mu$ is a line connecting three points with nonzero y-coordinates in the (x, y) plane, then dividing by μy and rearranging yields the line

$$s = \frac{-\lambda}{\mu} t + \frac{1}{\mu}$$

Then, the group law in the (t, s)-plane operates identically to the group law in the (x, y)-plane: one firsts draws a line connecting the two points one wishes to add (tangent if they are the same point). One then looks at the third intersection point (t_3, s_3), and draws a line connecting it and the identity element. In the (x, y) case, this was $\mathcal{O}$, but in our case, it is the origin. Then, $P_1 + P_2$ is equal to the third point on that line. In this case, however, this is simple: $(-t_3, -s_3)$ is both on the curve and the line connecting (t_3, s_3) and the origin.

One may be concerned about defining this group law when we have excluded the points of order 2. However, we will be restricting our focus on this group law to subsets (which will turn out to be subgroups) that do not contain these points.

The last necessary ingredient of our proof is the subring $R_p \subset \mathbb{Q}$ defined by

$$R_p = \{q \in \mathbb{Q} \mid \nu_p(q) \geqslant 0\}$$

Note that $0 \in R_p$. Oftentimes the subscript p will be omitted [9, p.49–51].

We are now equipped to prove

Proposition 4.2. *The set $C(p^\varepsilon)$ is a subgroup of $E(\mathbb{Q})$ for all $\varepsilon > 0$.*

Proof. Let $(x, y) \in C(p^\varepsilon)$. Then, writing

$$(x, y) = \left(\frac{n_x}{d_x p^{2(\varepsilon+i)}}, \frac{n_y}{d_y p^{3(\varepsilon+i)}} \right)$$

we see that

$$t = \frac{\frac{n_x}{d_x p^{2(\varepsilon+i)}}}{\frac{n_y}{d_y p^{3(\varepsilon+i)}}} = \frac{n_x d_y}{n_y d_x} p^{\varepsilon+i} \quad , \quad s = \frac{d_y}{n_y} p^{3(\varepsilon+i)}$$

This implies $(x, y) \in C(p^\varepsilon)$ iff $t \in p^\varepsilon R$. Broadly speaking, it suffices to work in the (t, s)-coordinates to prove $C(p^\varepsilon)$ is a subgroup. One would expect it easier to work in the (t, s)-plane because the sets we care about are neighborhoods of the origin.

Suppose $P_1, P_2 \in C(p^\varepsilon)$ are distinct points but $t_1 = t_2$. Then, the vertical line through them intersects the curve at some point (t_1, s_3), whose additive inverse is $(-t_1, -s_3)$. Hence, the t-coordinate of $P_1 + P_2$ is in $p^\varepsilon R$, and so $P_1 + P_2 \in C(p^\varepsilon)$ [10, p.51].

Otherwise, suppose $P_1, P_2 \in C(p^\varepsilon)$ are either equal or have distinct t-values.

When $t_1 \neq t_2$, there exists a line L given by $s = \alpha t + \beta$ through (t_1, s_1) and (t_2, s_2). Then, the t-values of the three points of intersection of L with the curve are roots of

$$\alpha t + \beta = t^3 + at^2(\alpha t + \beta) + bt(\alpha t + \beta)^2 + c(\alpha t + \beta)^3$$

Expanding yields

$$0 = (1 + a\alpha + b\alpha^2 + c\alpha^3)t^3 + (a\beta + 2b\beta\alpha + 3c\beta\alpha^2)t^2 + \dots$$

This polynomial gives the sum of its roots:

$$t_1 + t_2 + t_3 = \frac{a\beta + 2b\beta\alpha + 3c\beta\alpha^2}{1 + a\alpha + b\alpha^2 + c\alpha^3}$$

where t_3 is the third point of intersection of L with the curve [9, p.53].

This equation motivates finding an expression for $\alpha = (s_2 - s_1)/(t_2 - t_1)$. The (t_i, s_i)'s satisfy $s = t^3 + at^2s + bts^2 + cs^3$, and subtracting these expressions for s_1 and s_2 gives

$$s_2 - s_1 = (t_2^3 - t_1^3) + a\left[t_2^2 s_2 - t_1^2 s_1\right] + b\left[t_2 s_2^2 - t_1 s_1^2\right] + c(s_2^3 - s_1^3)$$

Clever addition and subtraction of $t_1^2 s_2$ and $t_1 s_2^2$, factoring, and use of the difference of squares and cubes formulae gives

$$\alpha = \frac{s_2 - s_1}{t_2 - t_1} = \frac{(t_2^2 + t_2 t_1 + t_1^2) + a(t_2 + t_1)s_2 + bs_2^2}{1 - at_1^2 - bt_1(s_2 + s_1) - c(s_2^2 + s_1 s_2 + s_1^2)}$$

What about when $(t_1, s_1) = (t_2, s_2)$? Letting $(t, s) = (t_1, s_1)$ and implicitly differentiating $s = t^3 + at^2s + bts^2 + cs^3$ gives

$$s' = 3t^2 + 2ats + at^2 s' + bs^2 + 2btss' + 3cs^2 s'$$

which simplifies to

$$s' = \frac{3t^2 + 2ats + bs^2}{1 - at^2 - 2bts - 3cs^2}$$

Note that this formula for the slope of the tangent line is identical to the previous one for α if we have $t_1 = t_2$ and $s_1 = s_2$, justifying the use of the previous expression for α in every case.

We now have all the ingredients we need to prove that $-t_3 \in p^\varepsilon R$. In the expression for α, the numerator and the denominator minus 1 is in $p^{2\varepsilon}R$ because each of t_1, t_2, s_1, s_2 is in $p^\varepsilon R$. Hence, the denominator is a unit and $\alpha \in p^{2\varepsilon}R$. Similarly, $\beta = s_1 - \alpha t_1$ is in $p^{3\varepsilon}R$. Note that both of these use the fact that $f(x) \in \mathbb{Z}[x]$. Lastly, $t_1 + t_2 + t_3 \in p^{3\varepsilon}R$ because the denominator is a unit and the numerator is in $p^{3\varepsilon R}$. Because $t_1, t_2 \in p^\varepsilon R$, we conclude $-t_3 \in p^\varepsilon R$.

We have shown that $C(p^\varepsilon)$ is a subgroup, as it is closed under taking negatives (clearly) and addition [9, p.54]. ■

Proposition 4.3. *The map*

$$\begin{aligned} t\colon C(p^\varepsilon)/C(p^{3\varepsilon}) &\hookrightarrow p^\varepsilon R/p^{3\varepsilon}R \\ (x, y) &\mapsto \frac{x}{y} \pmod{p^{3\varepsilon}R} \end{aligned}$$

is an injective homomorphism.

Note that these are really infinitely many maps indexed by p and ε, which have been fixed throughout our discussion.

Proof. In the proof of the previous proposition, we saw that the statement $t_1 + t_2 + t_3 \in p^{3\varepsilon}R$ held. This implies $t(P_1) + t(P_2) - t(P_1 + P_2) \in p^{3\varepsilon}R$. This is the condition for t being a homomorphism.

It remains to compute $\ker t$. If $t(P) = x/y \in p^{3\varepsilon}R$, then $P \in C(p^{3\varepsilon})$ by the correspondence between (x, y) and (t, s) mentioned earlier. The reverse inclusion holds as well: if $t(P) \in p^{3\varepsilon}R$, then $P \in C(p^{3\varepsilon})$ [9, p.54]. ■

Corollary 4.4. *$C(p) - \{\mathcal{O}\}$ contains no points of finite order.*

Proof. Suppose not. Let P be of order $m \neq 1, 2$, as the $m = 2$ case has already been addressed.

Suppose first that $p \nmid m$. Then, using that $P \neq \mathcal{O}$, choose the maximal $\varepsilon > 0$ such that P is in $C(p^{\varepsilon})$ but not in $C(p^{\varepsilon+1})$. The fact that t is a homomorphism gives

$$0 = t(\mathcal{O}) = t(mP) \equiv mt(P) \pmod{p^{3\varepsilon}R}$$

But m is a unit and hence $t(P) = m^{-1}0 = 0$. This means $t(P) \in p^{3\varepsilon}R$, and hence $P \in \ker t = C(p^{3\varepsilon})$. But if $3\varepsilon > \varepsilon + 1$ contradicts the maximality of ε.

Suppose now that $p \mid m$. Write $m = pn$. Note that nP is of order p and contained in $C(p)$ because it is a subgroup. Once again, choose the maximal $\varepsilon > 0$ such that $nP \in C(p^{\varepsilon})$ and $nP \notin C(p^{\varepsilon+1})$. Then,

$$0 = t(\mathcal{O}) = t(pnP) \equiv pt(nP) \pmod{p^{3\varepsilon}R}$$

This implies $t(nP) \in p^{3\varepsilon-1}R$. But $3\varepsilon - 1 \geqslant \varepsilon + 1$ contradicts the maximality of ε [9, p.55–6]. ■

To summarize, we have proved

Corollary 4.5. *For $y^2 = f(x)$ where $f \in \mathbb{Z}[x]$ is monic, $\mathrm{Tor}(E(\mathbb{Q})) \subset \mathbb{Z}^2$.*

The Nagell-Lutz theorem follows from this corollary and the next proposition:

Proposition 4.6. *Suppose $P = (x, y) \in \mathrm{Tor}(E(\mathbb{Q}))$ and $y \neq 0$. Then, $y^2 | D$.*

Proof. Note that because P is of finite order, so is $2P$. Hence both points have integer coordinates. The duplication formula is $x(2P) = g(x)/4y^2$, where g is a polynomial of degree 4. When a, b, and c are integers, $g \in \mathbb{Z}[x]$. Furthermore, note that the duplication formula tells us $4y^2x(2P) = g(x)$. All of these quantities are integers as $2P$ is also of finite order, so $y^2 | g(x)$.

Then, there exist polynomials $r, s \in \mathbb{Z}[x]$ such that $D = rf + sg$ [7]. Because these polynomials all have integer coefficients, they are all integers when evaluated at x. But y^2 divides $f(x)$ and $g(x)$, and hence $y^2|D$. ■

In fact, the Nagell-Lutz theorem was considerable strengthened by Mazur in 1975:

Theorem 4.7. *For an elliptic curve with coefficients in* $\mathbb{Q}$ *and nonzero torsion,* $\mathrm{Tor}(E(\mathbb{Q}))$ *is isomorphic to either* $\mathbb{Z}/n\mathbb{Z}$*, where* $n = 1, 2, \ldots, 10, 12$ *or* $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z}$*, where* $n = 2, 4, 6, 8$ *[7, p.68].*

Note that the surprising part of this theorem is the small number of sizes of $\mathrm{Tor}(E(\mathbb{Q}))$. The structure of these groups is not nearly as surprising, as it is a not-difficult fact which follows from topology that $E(\mathbb{R})$ is isomorphic to $\mathbb{R}/\mathbb{Z}$ or $\mathbb{R}/\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$ depending on the number of connected components [8]. Hence, the torsion subgroup of $E(\mathbb{Q})$ will be of the form described in the theorem, ignoring the size.

5 The rank of $E(\mathbb{Q})$

The Nagell-Lutz theorem gives a finite algorithm for finding the torsion points of elliptic curves in a specific form. One simply iterates over the various possible values of y, i.e. the values of y satisfying $y^2|D$, and then solves the resulting cubic equation using the cubic formula to find possible values of y. Note however that the Nagell-Lutz is not an if and only if statement, so any rational point satisfying the conditions may actually be of infinite order. However, the theorem from Mazur shows that torsion elements can only have order less than 13, so a rational point calculated to have order 13 or greater is necessarily of infinite order.

There is an—ultimately naïve—hope that there is an algorithm for finding the rank of an elliptic curve. No such algorithm, however, has been found, and finding the rank of an elliptic curve remains a difficult yet fruitful problem. Here are some remarks illustrating how the rank of an elliptic curve is a deceptively tricky and incredibly interesting problem, and how the numbers involved in finding the rank and rational points of an elliptic curve can get larger than is feasible for fast calculation:

- The rank of an elliptic curve is the subject of many open problems, perhaps most famously the Birch–Swinnerton-Dyer conjecture. This Millennium Prize Problem states that the rank of an elliptic curve over a number field is equal to the order of the zero of the *Hasse–Weil L-function* $L(s)$ of the elliptic curve at $s = 1$, where this order could be 0 if L does not vanish at $s = 1$. Prior to the proof of the "modularity theorem" in

[7] Once one is assured of the existence of these polynomials and their degrees (in this case, r is of degree 3 and s is of degree 2), finding these polynomials is simple but tedious. One first multiplies out $rf + sg$. Equating coefficients yields a 7×7 system of linear equations. This can then be solved with your favorite program (I used Mathematica) to get

$$r(x) = 3x^3 - ax^2 - 6bx + 2ab - 27c$$
$$s(x) = -3x^2 - 2ax + a^2 - 4b$$

the early 2000s, it was not even known whether $L(s)$ could be analytically extended past the line $\mathrm{Re}(s) > 3/2$, and even now it is only known that there is a continuation when the base field is $\mathbb{Q}$ [6, p.90–1] [4] [15].

- The Birch–Swinnerton-Dyer conjecture has an interesting corollary that bears mentioning. Tunnell's theorem, which assumes the conjecture, gives a criterion calculable in finite time for calculating if a number is a congruent numbers. Namely, let n be a squarefree number (as those positive integers with squares can quickly be reduced to the squarefree case). If n is odd, it is a congruent number if and only if

$$\#\left\{x, y, z \mid n = 2x^2 + y^2 + 32z^2\right\} = \frac{1}{2}\left\{x, y, z \in \mathbb{Z} \mid n = 2x^2 + y^2 + 8z^2\right\}$$

  Similarly, n even is a congruent number if and only if

$$\#\left\{x, y, z \mid n = 8x^2 + 2y^2 + 64z^2\right\} = \frac{1}{2}\left\{x, y, z \in \mathbb{Z} \mid n = 8x^2 + 2y^2 + 16z^2\right\}$$

  Hence, translating the congruent number into its equivalent elliptic curve formulation, as was done in the first section and which may have seemed gratuitous, has yielded a partial resolution to the problem [6, p.221].

- An upper bound on the possible rank of an elliptic curve has not been proven or disproven, though almost all elliptic curves found have small rank. Currently, the record for the rank of an elliptic curve was found by Elkies in 2006: the curve $y^2+xy+y = x^3-x^2-ax+b$, where a is 20 067 762 415 575 526 585 033 208 209 338 542 750 930 230 312 178 956 502 and b is 34 481 611 795 030 556 467 032 985 690 390 720 374 855 944 359 319 180 361 266 008 296 291 939 448 732 243 429, has rank $\geqslant 28$ [5].

- Rational points themselves can be more complicated (i.e. larger in height) than may be expected from the numbers occurring in $f(x)$. Bremner and Cassels showed the elliptic curve $y^2 = x^3 + 877x$, which has discriminant D on the order of 10^9, has $E(\mathbb{Q}) \simeq \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$. However, the torsion-free part of $E(\mathbb{Q})$ is generated by (x, y), where

$$x = \frac{3754945281271621931055040699420927923462 01}{6215987776871505425463220780697238044100}$$

  This means the non-logarithmic height of (x, y) is approximately 3.75×10^{41}, which is roughly $D^{4.41}$. They conclude that "unless one has strong reason to believe that a rational point exists, there is a marked reluctance to persevere in a search once the numbers cease to be small" [2].

However, using the tools of our proof of the Mordell theorem, we can still make some headway on finding the rank and rational points of an elliptic curve. As with the Mordell theorem, we will restrict our attention to curves $y^2 = x^3 + ax^2 + bx$. This first requires

understanding $[G : 2G]$. If $G \simeq \mathbb{Z}^r \oplus \mathbb{Z}/p_1^{\varepsilon_1}\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/p_s^{\varepsilon_s}\mathbb{Z}$, where $p_1, \ldots, p_s$ are not necessarily unique primes, then

$$\frac{G}{2G} \simeq \left(\frac{\mathbb{Z}}{2\mathbb{Z}}\right)^r \oplus \frac{\mathbb{Z}/p_1^{\varepsilon_1}\mathbb{Z}}{2\mathbb{Z}/p_1^{\varepsilon_1}\mathbb{Z}} \oplus \cdots \oplus \frac{\mathbb{Z}/p_s^{\varepsilon_s}\mathbb{Z}}{2\mathbb{Z}/p_s^{\varepsilon_s}\mathbb{Z}}$$

But

$$\frac{\mathbb{Z}/p_i^{\varepsilon_i}\mathbb{Z}}{2\mathbb{Z}/p_i^{\varepsilon_i}\mathbb{Z}} \simeq \begin{cases} \mathbb{Z}/2\mathbb{Z} & \text{if } p_i = 2 \\ 0 & \text{else} \end{cases}$$

It follows that

$$[G : 2G] = 2^{r + \#\{j \mid p_j = 2\}}$$

The second expression in the exponent can be found to equal

$$\#G_2 = \#\{g \in G | 2g = \mathcal{O}\}$$

by looking at which linear combinations of the generators of G are of order 2 [9, p.90]. $\#G_2$ is either 2 or 4 depending on whether f has 1 or 3 real roots (i.e. whether $a^2 - 4b$ is not or is a square).

On the other hand, using the maps φ and ψ from the proof of the Mordell theorem, note that $2G = \psi \circ \varphi(G)$. Hence,

$$[G : 2G] = [G : \psi(\overline{G})][\psi(\overline{G}) : \psi \circ \varphi(G)] = [G : \psi(\overline{G})]\frac{[\overline{G} : \varphi(G)]}{[\ker \psi : \varphi(G) \cap \ker \psi]}$$

where the last equality holds from basic isomorphism theorems [9, p.91]. We knowker $\psi = \{\overline{\mathcal{O}}, \overline{T}\}$ and $\overline{T} \in \varphi(G)$ iff $\overline{b}$ is a square, so the index in the denominator above is either 1 or 2, depending on if $\overline{b}$ is or is not a square, respectively.

Putting all this information together, we get

$$2^r = \frac{1}{4}[G : \psi(\overline{G})][\overline{G} : \varphi(G)]$$

Using the map α from earlier, we had that $G/\psi(\overline{G}) \simeq \alpha(G)$, and hence these indices can sometimes be computed with α and $\overline{\alpha}$ can be computed by looking at $\#\alpha(G)$ and $\#\overline{\alpha}(\overline{G})$ [9, p.93].

We must then acquaint ourselves with the image of α. We write (x, y) as $(m/e^2, n/e^3)$. If $x = 0$, then $(x, y) = T$ and $\alpha(T) = b$. Similarly, if $a^2 - 4b = w^2$ is a square, then G contains the points $((-a \pm w)/2, 0)$ by the quadratic formula.

Suppose now that $m, n \neq 0$. These points satisfy $n^2 = m(m^2 + ame^2 + be^4)$.

Writing $b_1 = \pm \gcd(m, b)$ where we choose the sign so $mb_1 > 0$, we have $m = b_1 m_1$ and $b = b_1 b_2$ where m_1 and b_2 are relatively prime and $m_1 > 0$. After substituting these values into the previous equation, we see that $b_1 | n$. Writing $n = b_1 n_1$ for some n_1 and substituting into the previous equation again yields $n_1^2 = m_1(b_1 m_1^2 + am_1 e^2 + b_2 e^4)$. Note that the two terms on the right hand side are relatively prime; otherwise, m_1 would share

a factor with b_2e^4 since m_1 shares its factors with the first two terms in the parentheses. This contradicts $\gcd(m_1, b_2) = \gcd(m_1, e) = 1$. Because the product of these two terms is a square, being relatively prime implies both terms themselves are squares. We write $m_1 = M^2$, $N^2 = b_1m_1^2 + am_1e^2 + b_2e^4$, and $n_1 = MN$. It follows that x and y can be written as

$$x = \frac{b_1M^2}{e^2} \quad , \quad y = \frac{b_1MN}{e^3}$$

Modulo squares, then, x is one of the finitely many divisors of b. Furthermore, substituting the numerators of x and y into the equation for n_1^2 before and eliminating m_1 gives $N^2 = b_1M^4 + aM^2e^2 + b_2e^4$. This is a condition on the admissible values of M, N, and e. There are some others as well: because x and y are in lowest terms, $\gcd(M, e) = \gcd(N, e) = \gcd(b_1, e) = 1$. Another condition is $\gcd(b_2, M) = 1$, as $M^2 = m_1$ and $\gcd(m_1, b_2) = 1$. The last condition is $\gcd(M, N) = 1$, as $\gcd(M^2, N^2) = 1$.

In summary, to find $\#\alpha(G)$, we write down the equation $N^2 = b_1M^4 + aM^2e^2 + b_2e^4$ for each pair of divisors b_1, b_2 and look for solutions that fit the aforementioned conditions.

For example, the earlier points of G, $((-a \pm w)/2, 0)$ are accounted for by this method, as $b = (-a + w)(-a - w)/4$, and $(M, N, e) = (1, 0, 1)$ solves the necessary equation [9, p.91–4].

The only problem with this method for computing rank—and it is a large one—is that there is no known method for finding solutions to the equation $N^2 = b_1M^4 + aM^2e^2 + b_2e^4$.

We illustrate this technique with an example. Consider C defined by $y^2 = x^3 - 5x$, noting that $\overline{C} = x^3 + 20x$. The possibilities for b_1 then are $\pm1, \pm5$. The four equations we will look at are

$$N^2 = \pm M^4 \mp 5e^4 \quad , \quad N^2 = \pm 5M^4 \mp e^4$$

The solution to the first two equations are $(M, N, e) = (3, 1, 2)$ and $(1, 2, 1)$. These are also the solutions to the second two equations. Using the formula for (x, y) as a function of b_1, M, N, and e, we get the rational points $(9/4, 3/8)$ and $(-1, -2)$. These solutions satisfying the relatively prime conditions we imposed. Hence, $\alpha(G) = \{\pm1, \pm5\}$, and so $\#\alpha(G) = 4$.

What about $\overline{\alpha}$? Since $\overline{b} = 20$, the possible values of $\overline{b}_1$ are $\pm1, \pm2, \pm4, \pm5, \pm10, \pm20$. Reducing these mod squares gives $\pm1, \pm2, \pm5, \pm10$. Note that the negative values will not work, as then $N^2 = \overline{b}_1M^4 + \overline{b}_2e^4$ has no nonzero real solutions. Note that 1 and 5 are in $\overline{\alpha}(\overline{G})$ as the images of $\overline{\mathcal{O}}$ and $\overline{T}$, respectively.

2 is not possible. If it were, $N^2 = 2M^4 + 10e^4$ would have a solution with $\gcd(M, 10) = 1$. By Fermat's Little Theorem, $M^4 \equiv 1 \mod 5$. Reducing the governing equation modulo 5 gives $N^2 \equiv 2 \mod 5$. But this has no solution. Hence, $2 \notin \alpha(\overline{G})$. It follows that neither is 10, as 5 is in the subgroup but 2 is not. We conclude that $\overline{\alpha}(\overline{G}) = \{1, 5\}$. Hence, the rank of $E(\mathbb{Q})$ is 1, as

$$2^r = \frac{\#\alpha(G)\#\overline{\alpha}(\overline{G})}{4} = 2$$

[9, p.94–5]. If one were to use the Nagell-Lutz criterion, they would find the rational points $(-1, \pm2)$ and $(5, \pm100)$.

6 Acknowledgments

I would like to thank my advisor, Dr. Bernd Siebert, for his guidance and patience with both me and my thesis. I would also like to thank my second reader, Dr. James Vick, for his continued support over my undergraduate career. Lastly, I would like to thank anyone who comforted me during this stressful process, whether they knew it or not.

7 Biography

Nathan Alvarez Olson was born in Austin in 1997. He enrolled in the Plan II Honors program and the Dean's Scholars program at the University of Texas at Austin in 2015. He majored in Plan II and Mathematics and received a certificate in Computer Science. During his time at UT, he studied abroad in Cuba, participated in an undergraduate REU at UC Berkeley, and helped middle and high school students learn math.

References

[1] Reza Akhtar. An introduction to elliptic curves. https://www.theoremoftheday.org/Docs/RezaAkhtar.pdf.

[2] A. Bremner and J. W. S. Cassels. On the equation $y^2 = x(x^2 + p)$. *Mathematics of Computation*, 42(165):257–264, 1984.

[3] Andries E. Brouwer. Mordell's theorem. https://www.win.tue.nl/~aeb/2WF02/mordell.pdf.

[4] Henri Darmon, Victor Rotger, and Yu Zhao. The Birch and Swinnerton-Dyer conjecture for $\mathbb{Q}$-curves and Oda's period relations. January 2012.

[5] Andrej Dujella. History of elliptic curves rank records. https://web.math.pmf.unizg.hr/~duje/tors/rankhist.html. [Online; accessed 15-April-2019].

[6] Neal I. Koblitz. *Introduction to Elliptic Curves and Modular Forms*. Graduate Texts in Mathematics. Springer-Verlag, 2 edition.

[7] J. S. Milne. *Elliptic Curves*. BookSurge Publishers, 2006.

[8] Joseph H Silverman. An introduction to the theory of elliptic curves. https://www.math.brown.edu/~jhs/Presentations/WyomingEllipticCurve.pdf.

[9] Joseph H. Silverman and John Torrence Tate. *Rational points on elliptic curves*. Undergraduate texts in mathematics. Springer-Verlag.

[10] Joseph H. Silverman and John Torrence Tate. *Rational points on elliptic curves*. Undergraduate texts in mathematics. Springer, 2nd edition.

[11] Karen E. Smith, Lauri Kahanpää, Pekka Kekäläinen, and William Traves. *An invitation to algebraic geometry*. Universitext. Springer, second edition.

[12] Michael Travis. Elliptic curves over $\mathbb{C}$. http://www.math.uchicago.edu/~may/VIGRE/VIGRE2008/REUPapers/Travis.pdf.

[13] Eric W. Weisstein. Elliptic curve. http://mathworld.wolfram.com/EllipticCurve.html.

[14] Eric W. Weisstein. Taniyama-Shimura Conjecture. http://mathworld.wolfram.com/Taniyama-ShimuraConjecture.html.

[15] Wikipedia contributors. Birch and Swinnerton-Dyer conjecture. https://en.wikipedia.org/wiki/Birch_and_Swinnerton-Dyer_conjecture, 2019. [Online; accessed 15-April-2019].