

Diophantine Equations and Elliptic Curves

Aadish Jain

Received May 17, 2024

Accepted July 21, 2024

Electronic access July 31, 2024

This paper will discuss the various methods and techniques used to look for integral solutions of equations. We cover the relatively elementary methods that can be used to solve equations of degree one in the full generality way possible and methods to generate more solutions to equations of degree two given a particular solution. Further, we introduce more advanced techniques that can be used to study equations of degree three. These include introducing elliptic curves and defining a group law on them. We further explain how to convert elliptic curves to the so-called Weirstrass normal form, which can be manipulated more straightforwardly. Finally, we demonstrate these techniques by finding positive solutions to a famous equation.

Introduction

Diophantine equations are equations, typically polynomial, where one seeks integer or natural solutions. Diophantine refers to Diophantus of Alexandria, one of the first mathematicians to introduce symbolism to algebra. Famous equations include:

$$x^n + y^n = z^n$$

Fermat famously conjectured this to have no positive integer solutions for $n > 2$, A conjecture which stood unproven for 358 years before being proven by Andrew Wiles in 1994¹. Another famous equation is:

$$x^3 + y^3 = z^3 + w^3$$

The smallest solution in distinct positive integers to this equation is $1^3 + 12^3 = 9^3 + 10^3 = 1729$. When the English mathematician G. H. Hardy remarked to the Indian number theorist Srinivasa Ramanujan that his taxi-cab number, 1729, seemed rather dull; Ramanujan replied that it was interesting, as 1729 is the smallest number that can be expressed as a sum of positive cubes in two different ways. However, while some Diophantine equations have been studied throughout history, the general case is impossible to solve. Hilbert's 10th problem² asked for a general algorithm that could determine whether a Diophantine polynomial has an integer solution, and 21 years of work by Martin Davis, Yuri Matiyasevich, Hilary Putnam, and Julia Robinson proved the DPRM theorem³, which shows such an algorithm cannot exist. Solving Diophantine equations of degree higher than two was a breakthrough in 20th-century mathematics. This paper will give an introduction to this theory and demonstrate it by finding positive integral solutions of the following equation:

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} = 4 \quad (1.1)$$

This equation was originally solved in a paper An unusual cubic representation problem⁴ which we will be using as a reference.

Section 2 discusses when and how equations of degree one can be solved for as many variables as possible. Section 3 discusses a connection between finding rational and integer solutions that will greatly benefit us in solving higher-degree equations. Section 4 explains how to solve homogeneous equations of degree two in three variables, covering some ideas that will help us later. In Section 5, we move to degree three and explain the group law of elliptic curves (equation (1.1) gives rise to a third-degree equation after multiplying by the common denominator) and their normal form and demonstrate both (the group law and the normal form) by solving the abovementioned equation.

Linear Diophantine Equation

Definition 2.1

A Linear Diophantine equation in n variables is an equation of the form:

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = c$$

with $c, a_1, a_2, \dots, a_n \in \mathbb{Z}$, with the coefficients a_i ($i = 1, 2, \dots, n$) being non-zero, and we are looking for integer solutions x_1, x_2, \dots, x_n .

Lemma 2.2

A linear Diophantine equation in n variables has a solution if and only if $(a_1, a_2, \dots, a_n) \mid c$, where (a_1, a_2, \dots, a_n) stands for the greatest common divisor of the a_i and \mid is the "divides" relation.

Proof. The result follows from the following generalization of Bézout's identity: for all non-zero integers a_1, a_2, \dots, a_n and any integer k , there exist $x_1, x_2, \dots, x_n \in \mathbb{Z}$ such that

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = k \cdot (a_1, a_2, \dots, a_n).$$

It is sufficient to show this for $k = 1$ as then:

$$\begin{aligned} & a_1x_1 + a_2x_2 + \dots + a_nx_n \\ & a_1x_1 + a_2x_2 + \dots + a_nx_n = (a_1, a_2, \dots, a_n) \\ & (ka_1) \cdot x_1 + (ka_2) \cdot x_2 + \dots + (ka_n) \cdot x_n \\ & = k \cdot (a_1, a_2, \dots, a_n) \end{aligned}$$

Now, consider the set:

$$S = \{m \in \mathbb{N} \mid m = a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n\}$$

This is the set of all positive linear combinations of the a_i 's. Since $a_1^2 + a_2^2 + \dots + a_n^2 \in S$, S is a non-empty subset of \mathbb{N} . This means that S has a least element (well-ordering principle), say d . We claim $d = (a_1, a_2, \dots, a_n)$. For any $q \in S$, by the division algorithm, we have $q = p \cdot d + r$ with $0 \leq r < d$. Since $r < d$ and d is the least element of S , $r \notin S$. But $r = q - p \cdot d$, so r is a linear combination of the a_i . But r is not in S , so it must be non-positive. Since it's also non-negative, it must be 0. Since the remainder on the division of q by d is 0, $d \mid q$. This holds for all such q , so d divides every element of S . Now note that for every a_i , one of the following holds:

$$\begin{aligned} a_i > 0 & \implies a_i \in S \implies d \mid a_i \\ a_i < 0 & \implies (-a_i) \in S \implies d \mid (-a_i) \implies d \mid a_i \end{aligned}$$

So d divides each of the a_i and hence $d \mid (a_1, a_2, \dots, a_n)$. But note d is a linear combination of the a_i which means that $(a_1, a_2, \dots, a_n) \mid d$. It follows that $d = (a_1, a_2, \dots, a_n)$.

Now, we demonstrate a method to solve a linear Diophantine equation in two variables by example.

Example 2.3

Find a solution to the $14x + 38y = -10$ over the integers.

First, note that $(14, 38) = 2$ and $2 \mid -10$. This shows that this equation has solutions. Now, we will use the Euclidean algorithm on 14 and 38:

$$\begin{aligned} 38 &= 14 \cdot 2 + 10 \\ 14 &= 10 \cdot 1 + 4 \\ 10 &= 4 \cdot 2 + 2 \\ 4 &= 2 \cdot 2 + 0 \end{aligned} \tag{2.1}$$

By rearranging and discarding the last equation:

$$38 - 14 \cdot 2 = 10 \tag{2.2}$$

$$14 - 10 \cdot 1 = 4 \tag{2.3}$$

$$10 - 4 \cdot 2 = 2 \tag{2.4}$$

Using (2.3) and (2.4) we obtain:

$$10 - (14 - 10 \cdot 1) \cdot 2 = 2$$

$$10 - 14 \cdot 2 + 10 \cdot 2 = 2$$

$$10 - 14 \cdot 2 + 10 \cdot 2 = 2$$

$$10 \cdot 3 - 14 \cdot 2 = 2$$

Using (2.2) and (2.6) we get:

$$(38 - 14 \cdot 2) \cdot 3 - 14 \cdot 2 = 2$$

$$38 \cdot 3 + 14 \cdot (-8) = 2$$

Now, we multiply this equation by -5 :

$$38 \cdot (-15) + 14 \cdot 40 = -10$$

Hence, $x = 40, y = -15$ is a solution.

While $(40, -15)$ is a solution to the previous equation, so is $(21, -8)$. Let us now find all the solutions, not just one of them.

Lemma 2.4.

If $ax + by = c$, $a, b, c, x, y \in \mathbb{Z}$ has (x_0, y_0) as a solution, then the family of all solutions is given by:

$$\left(x_0 - k \cdot \frac{b}{(a, b)}, y_0 + k \cdot \frac{a}{(a, b)} \right)$$

Proof. We note that all pairs of numbers of this form are solutions. We shall now show that all solutions are of this form. We know (x_0, y_0) is a solution and assume (x_1, y_1) is another solution. We have:

$$ax_0 + by_0 = c$$

$$ax_1 + by_1 = c$$

Subtracting these gives:

$$a \cdot (x_0 - x_1) + b \cdot (y_0 - y_1) = 0$$

Rearranging and dividing by (a, b) we obtain:

$$\frac{a}{(a, b)} \cdot (x_0 - x_1) = \frac{b}{(a, b)} \cdot (y_1 - y_0)$$

$$\frac{a}{(a, b)} \mid \frac{b}{(a, b)} \cdot (y_1 - y_0)$$

(2.1) But note that:

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

$$\frac{a}{(a,b)} \mid (y_1 - y_0)$$

$$y_1 - y_0 = k \cdot \frac{a}{(a,b)}$$

Hence, y_1 is of the required form, and by substituting it in the original equation, we can show that x_1 is also of the corresponding form.

Similarly, we can analyze linear equations in more variables, as illustrated by the following example:

Example 2.5.

Solve $3x + 9y + 5z = 7$ over the integers. Define $w = x + 3y$. Hence $3w + 5z = 7$. This is a linear equation in two variables, something we know how to solve. Solving this, we get:

We now solve the equation

$$w = 4 + 5k, z = -1 - 3k$$

Now, we can solve the equation $x + 3y = 4 + 5k$. Treating k as a constant, we get:

$$w = 4 + 5k - 3n, \quad y = n$$

Hence, we conclude that our family of solutions is:

$$(4 + 5k - 3n, n, -1 - 3k), \quad n, k \in \mathbb{Z}$$

Remark 2.6.

Note that when defining w , we use $w = x + 3y$ and not $w = 3x + 9y$. This is because, in the second case, w could only have been a multiple of 3. So, some solutions to the equation in z and w would not have corresponded to the solutions of the original equation.

Projectivization

Before we move on to higher-degree equations, it is essential to understand some introductory projective geometry.

Definition 3.1.

A homogeneous equation is an equation $f(x_1, x_2, \dots, x_n) = 0$ such that if (x_1, x_2, \dots, x_n) is a solution, then so is $(tx_1, tx_2, \dots, tx_n)$.

Example 3.2.

The equation $x^2 + y^2 = xz$ is homogeneous as:

$$x^2 + y^2 = xz \Rightarrow (tx)^2 + (ty)^2 = (tx) \cdot (tz)$$

Polynomial equations are homogeneous iff all (non-zero) terms in the equation have the same degree. Suppose we have an n -degree polynomial equation with integer coefficients in m variables as follows:

$$p(x_1, x_2, \dots, x_m) = 0 \quad (3.1)$$

Multiply all terms of this equation of degree $k < n$ by x_0^{n-k} to get a homogeneous equation:

$$q(x_0, x_1, x_2, \dots, x_m) = 0 \quad (3.2)$$

We call (3.1) the affine form and (3.2) the projective form. The process of converting an equation to a projective form is called homogenization. Since we will primarily be concerned with affine equations in two variables, we shall use the variable names z, x, y instead of x_0, x_1, x_2 for the remainder of this section.

Example 3.3.

A few examples of homogenization:

- The projective form of $x + 3y - 7 = 0$ is $x + 3y - 7z = 0$.
- The projective form of $x^2 + y^2 = 1$ is $x^2 + y^2 = z^2$.
- The projective form of $x^3 + 7y^2 + 9xy + 3x + 2 = 0$ is $x^3 + 7y^2z + 9xyz + 3xz + 2z^3 = 0$.

The inverse process, de-homogenization, is straightforward. Just substitute $z = 1$, and we get the affine form.

Since $x = y = z = 0$ is a solution to all projective equations, we call it a trivial solution and disregard it. Also, since multiplying a solution by a constant gives another solution, we consider all proportional solutions the same. For the remainder of this paper, whenever we talk of solutions of a projective equation, we mean non-zero solutions up to their multiples.

Now, we note the correspondence between the solutions of the projective and affine equations. We see that if (x_0, y_0) is a solution of the affine form, $(x_0, y_0, 1)$ is a solution of the projective form. Conversely, if (x_0, y_0, z_0) is a solution to the projective form, $(x_0/z_0, y_0/z_0)$ is a solution to the affine form, provided $z_0 \neq 0$. In particular, rational solutions to the affine equation correspond one-to-one, with a few exceptions, to integer solutions (after potentially multiplying by a common denominator) of the projective equation. This connection is precisely why these two forms will help solve Diophantine equations; it is often easier to convert a projective equation to affine form and look for rational solutions than to look for integer solutions directly. Now, we discuss what lines are in projective geometry and how they are related to the usual affine definition of a line.

Definition 3.4.

A projective point is any non-zero ordered triplet (x_0, y_0, z_0) defined up to its non-zero multiples.

Example 3.5.

$(1, 0, -1)$, $(-1, 0, 1)$, $(27, 0, -27)$ all represent the same projective point.

Definition 3.6.

A projective line is the set of all solutions to the projective equation $ax + by + cz = 0$ (for a, b, c not all 0). Further, we call three projective points collinear if they lie on a projective line.

Remark 3.7.

If a point A lies on a line l , it may be denoted as $A \in l$.

Remark 3.8.

Multiplying the equation of a projective line by any non-zero constant does not change it.

Theorem 3.9.

Projectivization preserves incidence relations, that is, if the projective point $A = (x_1, y_1, z_1)$ lies on the projective line $l : ax + by + cz = 0$ (with a and b , not both 0) then the affine point $A' = (\frac{x_1}{z_1}, \frac{y_1}{z_1})$ lies on the affine line $ax + by + c = 0$ if $z_1 \neq 0$. If instead $z_1 = 0$ then the affine line $ax + by + c = 0$ has slope $\frac{y_1}{x_1}$ (possibly infinite).

Proof.

$$A \in l \Rightarrow ax_1 + by_1 + cz_1 = 0$$

If $z_1 \neq 0$ we divide by it to obtain the required relation:

$$a \frac{x_1}{z_1} + b \frac{y_1}{z_1} + c = 0$$

If instead $z_1 = 0$ we get:

$$ax_1 + by_1 = 0 \Rightarrow \frac{-b}{a} = \frac{y_1}{x_1}$$

Since the slope of the line $ax + by = 0$ is $\frac{-b}{a}$, we are done.

Theorem 3.10.

Given any two distinct projective points, there is a unique projective line through both of them; given any two distinct projective lines, there is a unique projective point that lies on both.

Proof. Let the two points be $A : (x_1, y_1, z_1)$ and $B : (x_2, y_2, z_2)$. Consider the line:

$$l : (y_1z_2 - y_2z_1) \cdot x + (z_1x_2 - z_2x_1) \cdot y + (x_1y_2 - x_2y_1) \cdot z = 0$$

The coefficients of this equation are 0 iff $\frac{x_1}{x_2} = \frac{y_1}{y_2} = \frac{z_1}{z_2}$ which contradicts our assumption that A and B are distinct, so this equation does represent a line. We note that A and B lie on l . Suppose both A and B lie on another line $m : ax + by + cz = 0$. We will show that l and m are the same line. Since A and B lie on m , we have:

$$ax_1 + by_1 + cz_1 = 0 \quad \text{and} \quad ax_2 + by_2 + cz_2 = 0$$

We multiply the first equation by z_2 and the second by z_1 and subtract them from each other:

$$a \cdot (x_1z_2 - x_2z_1) + b \cdot (y_1z_2 - y_2z_1) = 0$$

Rearranging, we get:

$$\frac{y_1z_2 - y_2z_1}{a} = \frac{z_1x_2 - z_2x_1}{b}$$

Similarly, we can show that:

$$\frac{y_1z_2 - y_2z_1}{a} = \frac{z_1x_2 - z_2x_1}{b} = \frac{x_1y_2 - x_2y_1}{c}$$

Since the coefficients of l and m are proportional, they represent the same line.

Now consider 2 distinct lines $n : a_1x + b_1y + c_1z = 0$ and $o : a_2x + b_2y + c_2z = 0$. Consider the following point:

$$C : (b_1c_2 - b_2c_1, c_1a_2 - c_2a_1, a_1b_2 - a_2b_1)$$

By a similar argument as before, the coordinates of C are not all 0. So it is, in fact, a projective point. We see that it lies on both n and o . Now, suppose another point, say D , lies on both lines. We know that only a unique line can pass through two distinct points. But through C and D , two lines, n and o , pass. Hence, C must be the same point as D .

Remark 3.11.

Parallel lines in affine geometry intersect at a point with $z = 0$ in projective geometry.

Definition 3.12.

A line l is said to intersect a curve Ω with multiplicity n at a point P if substituting the equation of l in the equation of Ω allows us to factor $f(x, y, z)^n$ where $f(x, y, z) = 0$ is an equation of a line that intersects l at P (assuming the equation of l is not a factor of the equation of Ω).

Example 3.13.

Find the multiplicity of the intersection of the curve $x^2 + y^2 = z^2$ and the line $x = z$. Substituting the equation of the line in the curve, we get:

$$z^2 + y^2 = z^2 \implies y^2 = 0$$

Also, note that $x = z$ and $y = 0$ intersect at $(1, 0, 1)$. Since y^2 was our factor after the substitution, we say $x = z$ intersects $x^2 + y^2 = z^2$ at $(1, 0, 1)$ with multiplicity two.

Remark 3.14.

A way exists to define multiplicity for the intersection of any two general curves, but it is not required here. It has been described under the name ‘‘Intersection Numbers’’ in Section 3.3 of Algebraic Curves⁵.

Remark 3.15.

When considering intersections up to multiplicity, we mean that an intersection of multiplicity n is counted n times.

Theorem 3.16.

A projective line has at most n intersections with an n -degree projective curve unless the line is a factor of the curve.

Proof.This follows from substituting the equation of the line into the curve and using a similar argument as that for the proof that non-constant one-variable polynomials have at most n roots. The details are left to the reader.

Remark 3.17.

If we allow for complex numbers, then two projective curves of degree m and n with their equations having no common non-constant factor have exactly $m \cdot n$ intersections up to multiplicity. This result is Bézout’s Theorem, which can be found in Section 5.3 of Algebraic Curves⁵.

Definition 3.18.

A line L is said to be tangent to a curve Ω at the point P if P is the intersection point of L and Ω with a multiplicity of at least two.

Definition 3.19.

A point P is said to be an inflection point on a curve Ω if the tangent at P to Ω is a triple tangent; that is, it intersects Ω at P with multiplicity three.

Equations of Degree Two

Definition 4.1.

A three-variable homogeneous 2nd-degree Diophantine equation is an equation of the form:

$$Ax^2 + Bxy + Cy^2 + Dxz + Eyz + Fz^2 = 0$$

where $A, B, C, D, E, F \in \mathbb{Z}$. We will solve this equation for integers x and y . We start by de-homogenizing it by setting $z = 1$ and looking for rational points (for the solutions where $z = 0$, we can set either x or y to be 1). This equation is significantly more complicated than a linear equation, so let us start by analyzing a simple case:

Example 4.2.

Find all the rational solutions to $x^2 + y^2 = 1$, that is, find all rational points on the unit circle. Note that $(-1, 0)$ is a rational solution. A unique line with a real slope (non-vertical) joining them exists for any other point on this circle. Let this line be $y = tx + t$. By substituting this in the equation of the circle, we get:

$$\begin{aligned} x^2 + (tx + t)^2 &= 1 \\ (t^2 + 1) \cdot x^2 + 2t^2x + t^2 - 1 &= 0 \\ (x + 1) \cdot ((t^2 + 1) \cdot x + (t^2 - 1)) &= 0 \\ x = -1 \quad \text{or} \quad x &= \frac{1 - t^2}{1 + t^2} \\ y = tx + t & \end{aligned}$$

$$(x, y) = (-1, 0) \quad \text{or} \quad \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2} \right)$$

It is clear that there is a one-to-one correspondence between points on the circle (except $(-1, 0)$) and real values of t . We claim that there is a one-to-one correspondence between rational points on the circle (except $(-1, 0)$) and rational values of t . We can see that if $t \in \mathbb{Q}$, the corresponding point $\left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2} \right)$ will be a rational point. Conversely, if $\left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2} \right)$ is a rational point, then let $x = \frac{1 - t^2}{1 + t^2}, y = \frac{2t}{1 + t^2}$. Note that $x, y \in \mathbb{Q}$. Hence $t = \frac{y}{x + 1} \in \mathbb{Q}$ ($x + 1$ is non-zero). Hence, all rational points on the unit circle are of the form $\left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2} \right), t \in \mathbb{Q}$ or $(-1, 0)$. It is worth noting that $\lim_{t \rightarrow \infty} \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2} \right) = (-1, 0)$ which is to be expected as $t = \infty$ represents the vertical tangent to the unit circle at $(-1, 0)$ that is, a line whose ‘‘second’’ intersection (multiplicity two) with the unit circle is $(-1, 0)$, for the coming examples, we will assume that $t \in (-\infty, \infty]$.

The main idea from this example is that, given one rational point on a curve, we can use it to find the other rational points. We claim that the same method will work in the general case.

Theorem 4.3.

For any curve $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$ with $A, B, C, D, E, F \in \mathbb{Q}$ which has no non-constant factors and a rational point (x_0, y_0) on that curve, the set of all rational points on that curve is the set of second intersections (up to multiplicity) of lines with rational slope through (x_0, y_0) with that curve. We assume that vertical lines have a rational slope.

Proof. The line joining any two rational points must have a rational slope. We shall now prove that if a line with a rational slope passes through (x_0, y_0) , then its second intersection with the curve is rational. Let the line be:

$$y = tx + y_0 - tx_0, \quad t \in \mathbb{Q}$$

It suffices to prove that the x -coordinate of the second intersection is rational. Substitute the value of y in

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

and simplify to get:

$$P(t) \cdot x^2 + Q(t) \cdot x + R(t) = 0$$

$$P(t) = A + Bt + Ct^2$$

$$Q(t) = B(y_0 - tx_0) + 2C(y_0 - tx_0) + D + Et$$

$$R(t) = C \cdot (y_0 - tx_0)^2 + E \cdot (y_0 - tx_0) + F$$

But we know that $x = x_0$ satisfies this equation, and by Vieta's theorem, $x_0 + x_1 = -\frac{Q(t)}{P(t)}$. Hence, the x -coordinate of the second intersection is $-\frac{Q(t)}{P(t)} - x_0$, which is rational. (This works because $P(t) \cdot x^2 + Q(t) \cdot x + R(t)$ cannot be identically 0, because if it was, the equation would have the line as a factor, which we assumed was not possible).

Remark 4.4.

If our equation could be factored into two rational linear factors, we would have reduced it to linear equations, something we already know how to solve.

This theorem allows us to generate the entire family of solutions to equations of degree two given a particular solution.

Corollary 4.5.

If a line has $n - 1$ rational intersections, up to multiplicity, with an $n \geq 3$ degree rational curve, then it has an n th intersection that is also rational.

Proof. Since $n \geq 3$, the line has $n - 1 \geq 2$ rational points on it and hence has a rational slope. Now, we can substitute the equation of the line in the curve and use Vieta's theorem again to prove this result.

Remark 4.6.

The case where the line is a tangent can be checked by taking a derivative to find the tangent and noting that the derivative will be rational and the tangent will have an intersection with multiplicity two with the curve.

Let us use this theorem to find the rational points on a more complicated curve.

Example 4.7.

Find all rational points on the curve $y^2 + 3xy + x^2 + x + 7y = 0$. We note that $(0, 0)$ is a rational point on this curve. Consider the line $y = tx$ which passes through $(0, 0)$ for $t \in \mathbb{Q}$. By substituting the value of y in the curve, we obtain:

$$(tx)^2 + 3x \cdot tx + x^2 + x + 7tx = 0 \implies x \cdot ((t^2 + 3t + 1) \cdot x + 7t + 1) = 0$$

Hence, the second intersection of the line with the curve is on $x = \frac{-7t-1}{t^2+3t+1}$. Since it also lies on $y = tx$ we get that the set of all rational points on this curve is given by $\left(\frac{-7t-1}{t^2+3t+1}, \frac{-7t^2-t}{t^2+3t+1}\right)$ for all $t \in \mathbb{Q} \cup \{\infty\}$. In particular, note that if $t = \frac{-1}{7}$ this is $(0, 0)$ itself and if $t = \infty$ it is $(0, -7)$.

However, finding a particular solution may sometimes be difficult or even impossible. Consider the following example:

Example 4.8.

Find all rational points on the curve $x^2 + 6xy + 9y^2 = 7$.

$$x^2 + 6xy + 9y^2 = 7 \implies (x + 3y)^2 = 7$$

But $\pm\sqrt{7}$ is not rational. Hence, this curve has no rational point.

In general, it is always possible to determine if a given rational curve of degree two has a rational point in a finite number of steps. This follows from the Hasse-Minkowski theorem⁶. However, we shall not discuss it here. Further details can be found in Section 1.1 of *Rational Points on Elliptic Curves*⁷.

Equations of Degree Three and Elliptic Curves

For equations of degree one, we analyzed as general a case as possible. For equations of degree two, we restricted ourselves to homogeneous equations and analyzed them. For equations of degree three (the so-called cubic equations), we enter a territory that, while not uncharted, is very difficult to navigate. Here, we shall restrict ourselves to one particular problem, originally studied in the paper *An unusual cubic representation problem*⁴. The problem is to find positive integer solutions to:

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} = 4 \tag{5.1}$$

In this section, we will first discuss what a group is and how the set of solutions to this equation forms a group. Then, we shall convert this equation to a more straightforward form (the so-called Weierstrass normal form). Finally, we shall use the group law to generate rational solutions to the equation in the normal form until we get a positive solution.

Group Law

Similar to the 2nd-degree case, we wish for a way to generate more rational points on a 2nd-degree curve, given a rational point on it. Unfortunately, as the following example demonstrates, we can't use the same technique.

Example 5.1.

Find the rational solutions to $x^3 + y^3 = z^3$. We note that $(1, 0, 1)$ is a rational point on the curve. Consider a line $y = tx - tz$ for $t \in \mathbb{Q}$. Substituting it in our equation, we get:

$$x^3 + (tx - tz)^3 = z^3$$

$$(x - z) \cdot ((t^3 + 1)x^2 + (-2t^3 + 1)xz + (t^3 + 1)z^2) = 0$$

And here lies the problem: the second factor is of the second degree, so there is no guarantee that it has rational roots.

Remark 5.2.

By Fermat's Last Theorem¹, the only rational points on this curve are those where one of the variables is 0, that is, $(1, 0, 1)$, $(1, -1, 0)$, $(0, 1, 1)$. Hence, the only t for which the 2nd (and 3rd) intersections are rational is $t = -1$.

Instead, we will use (4.5) to define an operation on the set of rational points, which allows us to generate more points given any two. This operation will satisfy the group axioms.

Definition 5.3.

A set S with a binary operation $*$ is called a group iff:

1. $a * (b * c) = (a * b) * c$ for all $a, b, c \in S$ (Associativity).
2. There exists $e \in S$ such that for all $a \in S$, $a * e = e * a = a$ (Existence of an Identity element).
3. For all $a \in S$ there exists $b \in S$ such that $a * b = b * a = e$ (Existence of Inverses).

Further, if $a * b = b * a$ (Commutativity) for all $a, b \in S$, we say that the structure $(S, *)$ is an abelian group.

Example 5.4.

\mathbb{Z} with the addition operation is an abelian group, with identity 0. $\mathbb{Q} \setminus \{0\}$ is a group with operation multiplication and identity 1.

Now consider the projective curve:

$$\Phi : A_0x^3 + B_0x^2y + C_0xy^2 + D_0y^3 + E_0x^2z + F_0xyz + G_0y^2z + H_0xz^2 + I_0yz^2 + J_0z^3 = 0$$

$A_0, B_0, \dots, J_0 \in \mathbb{Z}$. Assume that the cubic is non-singular (every point has a unique tangent) and has no non-constant factors. Take the set of all rational points on Φ . Now consider an operation $*$ on this set such that $A * B$ is the third intersection of the line through A, B with Φ (if $A = B$, we consider the "3rd" intersection of the tangent at Φ). Note that:

$$A * B = C \iff A * C = B \iff B * A = C \quad (5.2)$$

This is because all three statements mean that a, b, c are collinear.

Definition 5.5.

We define addition on the rational points on Φ as $A + B = O * (A * B)$ for some fixed rational point O on Φ .

Theorem 5.6.

The set of all rational points on Φ with operation $+$ as defined above, is an abelian group.

Proof. Most requirements of being a group follow immediately; associativity is the only one that requires more work.

1. **Commutativity:** Since the line through A and B is the same as the line through B and A , we have:

$$A * B = B * A$$

$$O * (A * B) = O * (B * A)$$

$$A + B = B + A$$

2. **Identity:** We claim O is the required e . Using (5.2) with O, A and $O * A$, we have:

$$O * A = (O * A)$$

$$A = O * (O * A) = O + A = A + O$$

3. **Inverses:** For any given A , we claim that the corresponding B is $(O * O) * A$. First, we note that:

$$A + B = B + A = O * (A * B) = O * (A * ((O * O) * A))$$

Now, we use (5.2):

$$(O * O) * A = (O * O) * A$$

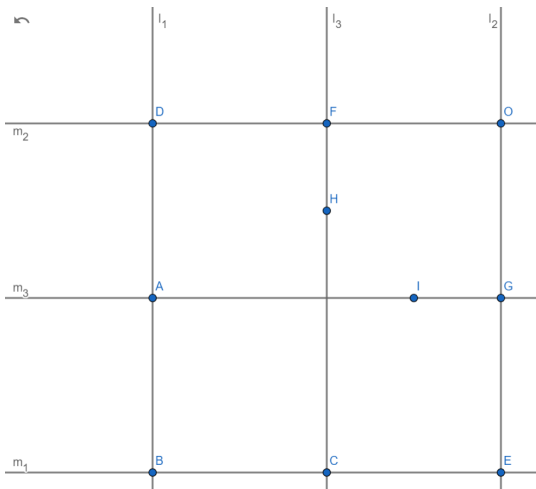
$$A * ((O * O) * A) = O * O$$

$$A + B = O * (O * O)$$

Using (5.2) again:

$$O * O = O * O \implies O * (O * O) = O$$

4. **Associativity:** Consider three points A, B, C on Φ and let $D = A * B$ and $E = B * C$. Let l_1 be the line through A, B and D , and let m_1 be the line through B, C and E . Let $F = A + B = O * D$ and $G = B + C = O * E$ and let m_2 be the line through O, D and F and l_2 be the line through O, E and G . Now, we assume that A, B, C, D, E, F, G, O are all pairwise distinct. The case where any of the points are equal has been skipped here; we invite the reader to fill in the details. Finally, consider $F * C = H$ and $A * G = I$. Let l_3 be the line through F, C and H , and m_3 be the line through A, G and I .



If we could prove that $H = I$, we would be done as then the following holds:

$$\begin{aligned} (A + B) + C &= F + C = O * (F * C) = O * H = O * I = O * (A * G) \\ &= A + G = A + (B + C) \end{aligned}$$

For the sake of contradiction, assume this was not the case. Now consider a cubic curve formed by multiplying the equation of l_1, l_2 and l_3 together $\Psi = l_1 \cdot l_2 \cdot l_3$ (the equation of Ψ is the product of the equations of l_1, l_2 and l_3) and similarly $\Omega = m_1 \cdot m_2 \cdot m_3$. A point lies on Ψ iff it lies on one of l_1, l_2 or l_3 and since all of them intersect Φ at three points each, which is the maximum (3.16) (We assumed that the equation of Φ has no non-constant factors), A, B, D, O, E, G, F, C and H are the only intersections of Ψ and Φ . Notably, I does not lie on Ψ ; similarly, H does not lie on Ω . Hence, Ψ and Ω are distinct

curves. Consider the family, say X , of cubic curves that are a linear combination of Ψ and Ω . That is, all the curves whose equations are of the form $a \cdot \Psi(x, y, z) + b \cdot \Omega(x, y, z) = 0$ where $\Psi(x, y, z) = 0$ is the cubic equation of Ψ and $\Omega(x, y, z) = 0$ is the equation of Ω . While we appear to have two degrees of freedom with this family, that of a and b , we only have one since multiplying by a constant does not change the curve. Since A, B, C, D, E, F, G and O lie on both Ψ and Ω , they lie on all curves in this family. Hence, this family is a subset of the family Y of all cubic curves through A, B, C, D, E, F, G and O (including the 0 curve, which has equation $0 = 0$ and passes through all projective points). We claim both families are the same. We will show this by showing that Y also has one degree of freedom. And if $X \subseteq Y$ and both have the same degrees of freedom, $X = Y$. An arbitrary cubic curve in three variables has ten coefficients, so it seems we would need ten equations to specify it. However, since multiplying by a constant does not change the curve, we only need nine. Since Y is a family of all curves that pass through eight distinct points, we already have eight equations; hence, only one more is needed to specify the curve completely. That is to say, the elements of Y have one degree of freedom. Clearly, $\Phi \in Y$ and hence $\Phi \in X$, so $\Phi(x, y, z) = a \cdot \Psi(x, y, z) + b \cdot \Omega(x, y, z)$. At I , Φ and Ω vanish and Ψ doesn't, so a must be 0. Similarly, b can be shown to be 0. But Φ is non-singular, so it can't be the 0 curve. So, we have a contradiction, which means our assumption is false. So $H = I$, and we are done.

Remark 5.7.

The argument for $X = Y$ via degrees of freedom can be formalized as follows: Y is a vector space of dimension two (here, we consider equations that differ by a constant multiple as different), and Ψ and Ω are linearly independent elements of Y . So Y is spanned by Ψ and Ω .

Remark 5.8.

An alternate method to prove associativity (and the rest of the group law) is deriving an explicit formula for $A + B$ and verifying the conditions using the formulas. This is not reasonable to do by hand, but software can do it for us. This can be used to check the special cases where two or more of A, B, C, D, E, F, G and O are the same.

Remark 5.9.

$$A + B + C = 0 \iff A, B, C \text{ are collinear.}$$

In the next section, we will be transforming between different forms of an equation, and each form will be a projective curve of degree three for which its group law can be defined. These group laws for each of these curves will be related in a way that preserves the structure of the group:

Definition 5.10.

If G is a group with operation $*$ and H is a group with operation $+$, then a function $f : G \rightarrow H$ is called a group homomorphism if for all $a, b \in G$, $f(a * b) = f(a) + f(b)$. That is, f preserves the group structure. Homomorphisms map the identity of G to the identity of H .

Example 5.11.

Consider the groups $(\mathbb{R}, +)$ and $(\mathbb{R}_{>0}, \cdot)$. The function $f : \mathbb{R} \rightarrow \mathbb{R}_{>0}$:

$$f(x) = e^x$$

is a homomorphism between these groups.

Weierstrass Form

Now that we have a group law, we multiply out the denominators to convert our equation to a form that is easier to work with.

$$a(a+b)(a+c) + b(b+c)(b+a) + c(c+a)(c+b) - 4(a+b)(b+c)(c+a) = 0 \tag{5.3}$$

We must be careful that this doesn't introduce any new solutions by multiplying by 0. Thankfully, that would require $a + b = 0$ or $b + c = 0$ or $c + a = 0$, which is impossible for positive integers.

Definition 5.12.

An elliptic curve is any projective curve of degree three in three variables with integer coefficients and at least one rational point on it.

Note that there exists a rational point $(a, b, c) = (-1, 0, 1)$ on (5.3). Hence, it is an elliptic curve. For our purposes, the most crucial fact about elliptic curves is that there exists a change of coordinates that takes any elliptic curve Φ with a rational point R_1 to the form:

$$\Psi : y^2z = x^3 + pxz^2 + qz^3, \quad p, q \in \mathbb{Q}$$

Such that the transformation is a group homomorphism between the group $(\Phi, +)$ and $(\Psi, +)$ with $+$ defined as in Definition 5.5 and “ O ” taken as $(0, 1, 0)$ on Ψ and some point R_2 on Φ (R_1 may or may not be the same as R_2). In general, such transformations are not simple. Say our transformation is f . We wish for f to be a bijection since we want to return to our original coordinates later. Since the identity element goes to the identity element, $f(R_2) = (0, 1, 0)$. Now consider the line $z = 0$ in the transformed coordinates. If we substitute it in n , we obtain that it intersects the curve thrice at $(0, 1, 0)$. That is, $(0, 1, 0)$ is an inflection point. But it may be that R_2 is not an inflection point. So, the pre-image of $z = 0$ can not be a line because no line intersects

Φ thrice at R_2 . So f could be a transformation that doesn't send lines to lines, and such transformations are more challenging to work with than ones that do.

Thankfully, we do not need to worry about this case as (5.2) has $(-1, 0, 1)$ as an inflection point, we will show that in such cases, a “line preserving” transformation exists that sends it to the normal form. Transformations that send lines to lines are of a specific form, as shown below (for the rest of this section, all variables are assumed to be rational).

Definition 5.13.

A projective transformation is a transformation from a coordinate system (a, b, c) to (x, y, z) of the following form:

$$l_{11}a + l_{12}b + l_{13}c = x$$

$$l_{21}a + l_{22}b + l_{23}c = y$$

$$l_{31}a + l_{32}b + l_{33}c = z$$

$$l_{11}l_{22}l_{33} + l_{12}l_{23}l_{31} + l_{12}l_{21}l_{32} - l_{11}l_{23}l_{32} - l_{12}l_{21}l_{33} - l_{13}l_{22}l_{31} \neq 0$$

Remark 5.14. This transformation may be written in a more enlightening way using matrices.

$$\begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

The last condition is there to ensure that the transformation is invertible since it corresponds to:

$$\left| \begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \right| \neq 0$$

This condition is equivalent to the three lines:

$$l_{11}a + l_{12}b + l_{13}c = 0$$

$$l_{21}a + l_{22}b + l_{23}c = 0$$

$$l_{31}a + l_{32}b + l_{33}c = 0$$

not being concurrent (all three not passing through a single projective point).

The inverse of a projective transformation is a projective transformation.

Remark 5.15.

Since they are invertible, a projective transformation may be stated by either expressing (x, y, z) in terms of (a, b, c) or vice versa.

Theorem 5.16.

A projective transformation sends lines to lines.

This theorem is demonstrated by the example below; generalizing this example involves quite a bit of algebra and has been skipped here. The reader is invited to fill in the details.

Example 5.17.

Consider the transformation:

$$x = a + b + c$$

$$y = a + c$$

$$z = a + b$$

Consider the set of all points on the lines $k_1 : a + b + c = 0$ and on $m_1 : x = 0$. Since $x = a + b + c$, any points that lie on k_1 have their images on m_1 , so the line k_1 is mapped to the line m_1 . Now consider the line $k_2 : b + c = 0$. Notice that:

$$b + c = 2(a + b + c) - (a + c) - (a + b)$$

So k_2 is mapped to $m_2 : 2x - y - z = 0$. Now consider the general case $k_3 : l_1a + l_2b + l_3c = 0$. We wish to express this in terms of x, y, z . For this, we will note that the inverse of our transformation is given by:

$$a = y + z - x$$

$$b = x - y$$

$$c = x - z$$

We substitute these in the equation of k_3 and get that k_3 maps to the line $l_1(y + z - x) + l_2(x - y) + l_3(x - z) = 0$.

Remark 5.18.

It is worth noting that a projective transformation is almost entirely specified by the pre-images of $x = 0, y = 0$ and $z = 0$. We have further choices with the scale; say the pre-image of $x = 0$ is $3a + 2b = 0$, then x may be equal to $k(3a + 2b)$ for any “scale factor” k , but since scaling all three pre-images means scaling our projective point, which doesn’t change it, we can always fix one of our scale factors as unity. This means that while we have nine coefficients in a projective transformation, we only have eight degrees of freedom.

It is clear that this result can be generalized. That is:

Corollary 5.19.

Projective transformations send curves of degree n to curves of degree n .

Another simple fact that follows from projective transformations being bijections is that incidence relations are preserved.

Lemma 5.20.

A point P lies on a curve Ω if $t(P)$ lies on $t(\Omega)$ where t is a projective transformation.

Proof. This fact follows from substituting the equations defining t in the equation of Ω . \square

Corollary 5.21.

Two projective curves Ω and Γ intersect at a point P if and only if $t(\Omega)$ intersects $t(\Gamma)$ at $t(P)$ where t is a projective transformation.

Remark 5.22.

Not only do projective transformations preserve intersections, but they also preserve the multiplicity of intersections. This has been stated in Section 5.1 of Algebraic Curves⁵.

Proof. Ω and Γ intersect at a point P if and only if P lies on both Ω and Γ . P lies on Ω if and only if $t(P)$ lies on $t(\Omega)$ and similarly for Γ . So P lies on both Ω and Γ if and only if $t(P)$ lies on both $t(\Omega)$ and $t(\Gamma)$ which happens if and only if $t(\Omega)$ and $t(\Gamma)$ intersect at $t(P)$. \square

Armed with these facts about projective transformations, we are ready to search for a transformation that sends an elliptic curve with an inflection point to the normal form. Let Ω be an elliptic curve and P be a rational point on it, which is also an inflection point. Let l be the triple tangent line to Ω at P . We wish for our transformation to send P to $(0, 1, 0)$. Since $(0, 1, 0)$ lies on $x = 0$ and $z = 0$, then P must lie on their pre-images. And since $z = 0$ is a triple tangent to the normal form at $(0, 1, 0)$, its pre-image must also be a triple tangent to Ω at P . Hence, the pre-image of $z = 0$ is l , the pre-image of $x = 0$ is some line through P except l , and since the three lines can’t intersect at a single point, the pre-image of $y = 0$ is some line not through P .

In fact, we don’t need any more conditions on the pre-images of $x = 0$ and $y = 0$. Only some transformations reduce the equation to the normal form. Still, any transformation such that the pre-images satisfy the given conditions will significantly simplify the equation of the elliptic curve to a point where a few simple transformations can finish the job. Suppose after performing one such transformation Ω goes to:

$$\Omega' : A_0x^3 + B_0x^2y + C_0xy^2 + D_0y^3 + E_0x^2z + F_0xyz + G_0y^2z + H_0xz^2 + I_0yz^2 + J_0z^3 = 0$$

By the nature of our transformations, $z = 0$ is a triple tangent at $(0, 1, 0)$. Substituting $z = 0$ we get:

$$A_0x^3 + B_0x^2y + C_0xy^2 + D_0y^3 = 0$$

Since we can factor out x^3 ($x = 0$ is the only line with no “ z ” term that passes through $(0, 1, 0)$), $B_0 = C_0 = D_0 = 0$. So Ω' is

of the form:

$$A_0x^3 + E_0x^2z + F_0xyz + G_0y^2z + H_0xz^2 + I_0yz^2 + J_0z^3 = 0$$

Now, consider the further transformation:

$$x = x' - \frac{E_0 \cdot z'}{3 \cdot A_0}$$

$$y = y' - \frac{F_0 \cdot x' + G_0 \cdot z'}{2 \cdot G_0}$$

$$z = z'$$

This removes the x^2z , xyz and y^2z terms, and we get our curve as:

$$\Omega'' : A_1 \cdot x'^3 + B_1 \cdot x'z'^2 + C_1 \cdot z'^3 + D_1 \cdot y'^2z' = 0$$

The only thing left to do is to make the coefficient of x'^3 and y'^2z' as 1, which can be achieved by the final transformation:

$$x' = \frac{D_1 \cdot x''}{A_1}$$

$$y' = \frac{D_1 \cdot y''}{A_1}$$

$$z' = z''$$

Composing the three transformations and renaming the variables, we obtain a transformation that takes Ω to the normal form. Note that the final transformation still takes P to $(0, 1, 0)$ and l to $z = 0$. Since the final transformation is a composition of projective transformations, which all take lines to lines and preserve incidence relations and tangents, and our group law was defined using incidence relations alone, it is clear that this is a homomorphism between the groups, $(\Omega, +)$ (technically, the set of rational points on Ω) with identity element as P and $(\Omega'', +)$ with identity $(0, 1, 0)$.

Remark 5.23.

A general method for converting curves to normal form without any known inflection point has been described in Section 1.3 of Rational Points on Elliptic Curves⁷ Now, let us perform these transformations on (5.2). We have the inflection point $(-1, 0, 1)$ and differentiating shows that the tangent to the curve at $(-1, 0, 1)$ is $6a - b + 6c = 0$. This is the pre-image of $z = 0$. We take the pre-image $x = 0$ to be $a + c = 0$ and that of $y = 0$ to be $a + b = 0$. So, our transformations are:

$$x = a + c$$

$$y = a + b$$

$$z = 6a - b + 6c$$

The inverse transformation here is:

$$a = -6x + y + z$$

$$b = 6x - z$$

$$c = 7x - y - z$$

We substitute this in (5.3) to get:

$$91x^3 - 27x^2z - 13xyz + 2xz^2 + y^2z + 2yz^2 = 0 \tag{5.4}$$

Now, we do a partial version of the next step. We eliminate the xyz and yz^2 terms, but not the x^2 term. This is done for no reason other than the fact that removing the x^2 term makes our equation more complicated instead of less, and there is nothing to be gained by removing the term. Our transformations are:

$$x = x'$$

$$y = y' - \frac{2z - 13x}{2}$$

$$z = z'$$

Substituting these in (5.4), we get:

$$91x'^3 - \frac{277}{4}x'^2z' + y'^2z' + 15x'z'^2 - z'^3 = 0 \tag{5.5}$$

We should ideally get rid of the 91 here, but once again, that will make stuff more complicated for no gain, so we instead get rid of just the denominators by the transformation:

$$x' = -2x''$$

$$y' = y''$$

$$z' = z''$$

Substituting this in (5.5) and rearranging, we get:

$$y''^2z'' = 728x''^3 + 277x''^2z'' + 30x''z''^2 + z''^3$$

Finally, we rename x'', y'', z'' as x, y, z .

$$y^2z = 728x^3 + 277x^2z + 30xz^2 + z^3$$

The conversion formulas between (5.3) and (5.6) are:

$$a = -x + y$$

$$b = -12x - z$$

$$c = -x - y$$

$$x = \frac{-(a+c)}{2}$$

$$y = \frac{a-c}{2}$$

$$z = 6a - b + 6c \tag{5.7}$$

The original paper⁴ on this equation chose a different “normal” form than ours. Instead of making the x^2z term 0, they made the z^3 term 0. Consider the following transformation:

$$x = \frac{x'}{728} - \frac{z'}{13}$$

$$y = \frac{y'}{728}$$

$$z' = z$$

Applying this on (5.6) and clearing the denominators gives:

$$y^2z = x^3 + 109x^2z + 224xz^2$$

This is the form used in that paper⁴ after substituting $N = 4$.

It is worth noting here that normal forms are generally not unique. We can always use a non-projective transformation to obtain a normal form, but even if we don't, we can choose which inflection point to send to $(0, 1, 0)$. For our equation, we could have used $(1, -1, 0)$ or $(0, 1, -1)$ instead of $(-1, 0, 1)$ just as well. We don't have uniqueness even if we fix the point we send to $(0, 1, 0)$ and restrict ourselves to projective transformations. We could always compose it with the further transformation:

$$x' = u^2 \cdot x$$

$$y' = u^3 \cdot y$$

$$z' = z$$

This would give a new transformation. However, this new transformation still sends the same lines to $x = 0$, $y = 0$, and $z = 0$. It just scales them differently. If we do not care about scale, then normal forms are unique. Formally:

Theorem 5.24.

$$\Gamma : A_0a^3 + B_0a^2b + C_0ab^2 + D_0b^3 + E_0a^2c + F_0abc + G_0b^2c + H_0ac^2 + I_0bc^2 + J_0c^3 = 0$$

If Γ is an elliptic curve with an inflection point $Q = (a_0, b_0, c_0)$ and there exist projective transformations f and g which convert from coordinates (a, b, c) to (x, y, z) and (x', y', z') respectively and send Γ to some normal forms, such that $f(Q) = (0, 1, 0)$ and $g(Q) = (0, 1, 0)$. Then, the pre-images of $x = 0, y = 0$ and $z = 0$.

with respect to f are the same as the pre-images of $x' = 0, y' = 0$, and $z' = 0$ with respect to g .

Proof. Let f^{-1} and g^{-1} be the inverse transformations of f and g .

$$m_1 : x = 0$$

$$m_2 : y = 0$$

$$m_3 : z = 0$$

$$n_1 : x' = 0$$

$$n_2 : y' = 0$$

$$n_3 : z' = 0$$

We wish to show that:

$$f^{-1}(m_i) = g^{-1}(n_i)$$

for $i = 1, 2$, and 3 .

Since $z = 0$ and $z' = 0$ must be triple tangents through $(0, 1, 0)$ in their respective coordinates, their pre-images must both be the triple tangent through Q . So $f^{-1}(m_3) = g^{-1}(n_3)$. Now consider the transformation $g \circ f^{-1}$ which converts from coordinates (x, y, z) to (x', y', z') and hence converts between the two normal forms. We wish to show that $g \circ f^{-1}(m_1) = n_1$ and $g \circ f^{-1}(m_2) = n_2$ because this would mean that $f^{-1}(m_1) = g^{-1}(n_1)$ and $f^{-1}(m_2) = g^{-1}(n_2)$ which is precisely what we wish to show. Let $g \circ f^{-1}$ be:

$$l_{11}x' + l_{12}y' + l_{13}z' = x$$

$$l_{21}x' + l_{22}y' + l_{23}z' = y$$

$$l_{31}x' + l_{32}y' + l_{33}z' = z$$

What we wish to show is equivalent to $l_{12} = l_{13} = l_{21} = l_{23} = l_{31} = l_{32} = 0$. Since we have already shown $f^{-1}(m_3) = g^{-1}(n_3) \iff g \circ f^{-1}(m_3) = n_3$, which means that $l_{31} = l_{32} = 0$. Now let the normal form in (x, y, z) coordinates be:

$$y^2z = x^3 + axz^2 + bz^3$$

After the transformation, it becomes:

$$(l_{21}x' + l_{22}y' + l_{23}z')^2(l_{33}z') = (l_{11}x' + l_{12}y' + l_{13}z')^3 + a(l_{11}x' + l_{12}y' + l_{13}z')(l_{33}z')^2 + b(l_{33}z')^3$$

Since this is a normal form, the coefficient of x'^3 is non-zero, and that of x'^2y' is 0. That is $l_{11}^3 \neq 0 \implies l_{11} \neq 0$ and $3l_{11}^2l_{12} = 0 \implies l_{12} = 0$. We now have:

$$(l_{21}x' + l_{22}y' + l_{23}z')^2(l_{33}z') = (l_{11}x' + l_{13}z')^3 + a(l_{11}x' + l_{13}z')(l_{33}z')^2 + b(l_{33}z')^3$$

This time, consider that the coefficient of y'^2z' is nonzero, and that of $x'y'z'$ is 0. So $l_{22}^2l_{33} \neq 0 \implies l_{22}, l_{33} \neq 0$ and $2l_{21}l_{22}l_{33} = 0 \implies l_{21} = 0$. We now have:

$$(l_{22}y' + l_{23}z')^2(l_{33}z') = (l_{11}x' + l_{13}z')^3 + a(l_{11}x' + l_{13}z')(l_{33}z')^2 + b(l_{33}z')^3$$

Now consider that the coefficient of x'^2z' and $y'z'^2$ must be 0 which leads to $l_{13} = 0$ and $l_{23} = 0$ respectively. And hence, we are done.

Remark 5.25

. The coefficients of x^3 and y^2z are the same in the normal form (the equation may be multiplied by any required rational to make them 1). So, we have $l_{22}^2 l_{33} = l_{11}^3$. Since projective points do not change on multiplication by a constant, neither do projective transformations. So, we may force $l_{33} = 1$ from which it can be shown to give $l_{11} = u^2$ and $l_{22} = u^3$ for some $u \in \mathbb{Q}$.

Finding a positive solution to the equation

We are now ready to start with our hunt for a positive solution to the equation:

$$a(a+b)(a+c) + b(b+c)(b+a) + c(c+a)(c+b) - 4(a+b)(b+c)(c+a) = 0 \tag{5.8}$$

We will start with a rational point A of (5.6) and compute $2A, 3A, \dots, nA$ where $nA = A + A + \dots + A$ with n summands. We will then convert these back to points on (5.8) and see if a, b, c are all positive or all negative. In either case, we can multiply by a common denominator and potentially -1 so that all three coordinates become positive integers, which will be our solution.

The only thing left to do is pick an A . We clearly cannot start from $(0, 1, 0)$ as $(0, 1, 0) + (0, 1, 0) = (0, 1, 0)$ because it is the identity. It is tempting to instead start with $(a, b, c) = (-1, 1, 0)$ or $(a, b, c) = (0, -1, 1)$. But since both of these are inflection points, so are their images, say A and B . Since A and O are both inflection points:

$$A * A = A \Rightarrow A + A = O * A$$

$$A * (A + A) = A * (O * A)$$

$$O * A = O * A \Rightarrow A * (O * A) = O$$

$$A * (A + A) = O \Rightarrow A + A + A = O * O = O$$

So $3A = O \Rightarrow 4A = A$, which means we are in a loop. This means we can't use this process recursively if we start from an inflection point. Another idea is to add A and B to get a new point. However, we note that $A + B = O$; hence, this idea also fails us. We have no choice but to find a new rational point on (5.8). The original paper⁴ on this equation found both $(11, 4, -1)$ and $(11, -9, 5)$ as other rational points and used the former to start generating points; we chose to use the latter. $(a, b, c) = (11, -9, 5)$ and its image after (5.7) $S_0 : (x, y, z) = (-10, 1, 125)$ will be our starting points.

Now we can start our search. We de-homogenize (convert to affine form) to simplify algebra by setting $z = 1$.

$$y^2 = 728x^3 + 277x^2 + 30x + 1 \tag{5.9}$$

Since this removes O from our curve, we must redefine our addition law. We have $A + B = O * (A * B)$. It follows from (3.9) $A * B$ can still be defined as the 3rd intersection of the line through A and B with (5.9) assuming that $A + B \neq O$ (we do not need to worry about this case) and $O * (A * B)$ is the reflection of $A * B$ across the x-axis. After de-homogenizing, our starting point becomes $S : (\frac{-2}{25}, \frac{1}{125})$.

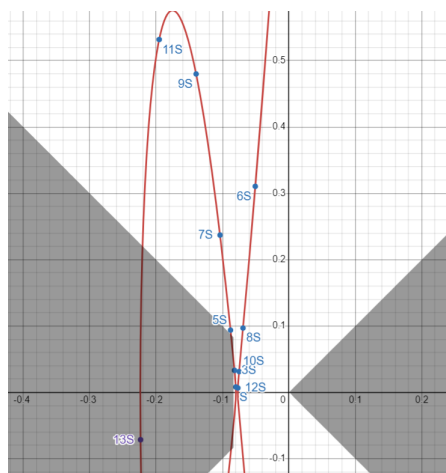
We will find $2S$ manually, this requires drawing a tangent to (5.9) at S , which can be obtained by differentiating (implicit) (5.9). The tangent comes out to be $y = -\frac{107x}{5} - \frac{213}{125}$. Substituting it back, we get the third intersection as $(\frac{143}{350}, \frac{-18283}{1750})$. Reflecting it about the x-axis, we get $2S = (\frac{143}{350}, \frac{18283}{1750})$. This corresponds to $(a, b, c) = (8784, -5165, -9499)$ on our original curve, not a positive solution.

Beyond this, we will use formulas to compute the remaining points. Since we will repeatedly add S , we derive the formula for $S + X$ where $X = (x_0, y_0)$. The line through S and X is given by $y = mx + c$ where $m = \frac{125y_0 - 1}{125x_0 + 10}$ and $c = \frac{x_0 + 10y_0}{125x_0 + 10}$. Substituting this in (5.9), we get:

$$(mx + c)^2 = 728x^3 + 277x^2 + 30x + 1$$

and by applying Vieta's theorem to obtain the sum of roots and noting that two of the roots are x_0 and $\frac{-2}{125}$ we get the third root to be, say $x_1 = \frac{2}{25} - x_0 + \frac{-277 + m^2}{728}$. Further, we have $y_1 = mx_1 + c$. Reflecting this across the x-axis, we get that $S + X$ is the point $(x_1, -y_1)$.

Now, we let the machine do the rest. It is worth graphing (5.9) to see how the positive solutions of (5.8) look.



The red curve is (5.9), the grey region corresponds to positive solutions of (5.8), and the points nS have been plotted in blue (4S lies outside the image, 13S has been plotted in purple). We see 13S lies in the grey region, so it is our positive solution. Converting it using (5.7) and multiplying out by the common

denominator to obtain our required solution (a_0, b_0, c_0) where:

$$a_0 = 1666647686543844986584613109531353154 \\ 0647604679654766832109616387367203990 \\ 6427643422481005348075794938744539548 \\ 5492535273990005122093641997167187559 \\ 4417036870073291371$$
$$b_0 = 1843865146707232952199146666910380962 \\ 7503176533640434051668643025780389550 \\ 6237580602582859039981257570380161221 \\ 6623981537942908215690451823856034188 \\ 67509209632768359835$$
$$c_0 = 3234342115382559235388065528522426333 \\ 0451946573450847101645239147091638517 \\ 6512509402068536126067685441814153553 \\ 5213607732730027180612906383302538977 \\ 2729796460799697289$$

We leave it to the reader to verify by hand that this is a solution.

Further Questions

The reader is undoubtedly left with many questions, so we will try to summarize them here.

1. Is this the smallest solution to (5.1)?
It is not so; the solution found in the original paper⁴ was smaller. In fact, they found the smallest positive solution. To prove that their solution is the smallest, knowledge of the rank of elliptic curves⁷ is needed, which is beyond the scope of this paper.
2. When can we find a positive solution to equations like (5.1) with 4 replaced by some N , and how does their size vary?
This question is more challenging to answer; the original paper⁴ managed to find some conditions of N ; in particular, N can't be odd, but finding necessary and sufficient conditions is more challenging. They also computed the sizes of positive solutions for $N \leq 200$.
3. When does a curve of degree three have a rational point on it?
No known algorithm can determine if a curve of degree three has a rational point in a finite number of steps⁷.
4. How does one solve Diophantine equations of degree higher than three?
Beyond the third degree lies the deep ocean, where particular cases like Fermat's Last Theorem¹ have been studied, but the general is unknown.

5. What happens if we introduce more variables?

This is also a dark forest, where little is known beyond the simple low-degree cases.

It seems incredible that such a simple-seeming question about integral points on curves has seemingly so little known about it.

Conclusion

This paper discusses the various methods used to solve Diophantine equations and explores equations of degrees one, two, and three. The theory of equations of degree one is fully explained here, and common ideas for degree two are also explored. We also briefly introduce degree three using tools from projective geometry, group theory, and the study of elliptic curves. We finally demonstrate these tools by solving an example equation.

References

- 1 A. Wiles, *Annals of Mathematics*, 1995, **141**, 443–551.
- 2 D. Hilbert *et al.*, *Bulletin-American Mathematical Society*, 2000, **37**, 407–436.
- 3 J. V. MATIJASEVIČ, in *ENUMERABLE SETS ARE DIOPHANTINE*, pp. 269–273.
- 4 A. Bremner and A. Macleod, *Annales Mathematicae et Informaticae*, 2014, **43**, 29–41.
- 5 W. Fulton, *Algebraic Curves: An Introduction to Algebraic Geometry*, Addison-Wesley Publishing Company, Advanced Book Program, 1989.
- 6 H. Hasse, *Journal für die reine und angewandte Mathematik*, 1923, **152**, 205–224.
- 7 J. Silverman and J. Tate, *Rational Points on Elliptic Curves*, Springer-Verlag, 1992.