SOLVING A DIOPHANTINE EQUATION WITH ELLIPTIC CURVES

DARREN YAO

Abstract. In this paper, we discuss the existence of positive integer solutions to $N = \frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b}$ and consider a method for finding integer solutions, when feasible. Much of the material comes from Brenmer and MacLeod's paper [1], but the $N = 6$ example and the computational complexity analysis are my original contributions.

1. Introduction

Diophantine equations are multivariate polynomial equations that are to be solved over the integers. Certain Diophantine equations can be solved with the use of substitutions to transform the equation into an elliptic curve of the form $y^2 = x^3 + ax + b$, no longer over the integers, but rather the rationals. This often allows us to use special properties of the elliptic curve to form a general solution to the original equation. In this paper, we will discuss how integer solutions to $N = \frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b}$ are related to the order of points on an elliptic curve. Furthermore, we will outline a method for finding solutions and discover that despite the apparent simplicity of the equation, minimal solutions can be extremely large, if they exist at all. Lastly, we will determine the feasibility of finding such solutions as the value of N grows larger.

Equation 1.1. *The equation we will discuss in this paper is* $N = \frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b}$*, where* N *is a positive integer.*

2. Finding Solutions

We will assume from now on that $N > 2$, because $N = 1$ has no solution (easily proven using AM-GM inequality) and the only solutions to $N = 2$ are multiples and/or permutations of the trivial solution $(a, b, c) = (1, 1, 3)$. Multiplying out the denominators, we see that the above equation can be converted to the third-degree Diophantine equation

$$N(a+b)(b+c)(c+a) = a(a+b)(c+a) + b(b+c)(a+b) + c(c+a)(b+c).$$

Applying the substitutions

$$x = \frac{-4(a+b+2c)(N+3)}{(2a+2b-c)+(a+b)N}, \ y = \frac{4(a-b)(N+3)(2N+5)}{(2a+2b-c)+(a+b)N}$$

transforms this equation into the elliptic curve

$$E : y^2 = x^3 + (4N^2 + 12N - 3)x^2 + 32(N+3)x.$$

The corresponding inverse transformations are

$$\frac{a}{a+b+c} = \frac{8(N+3)-x+y}{2(4-x)(N+3)}, \ \frac{b}{a+b+c} = \frac{8(N+3)-x-y}{2(4-x)(N+3)}, \ \frac{c}{a+b+c} = \frac{-4(N+3)-(N+2)x}{(4-x)(N+3)}.$$

These transformations are in arbitrary order; since Equation 1.1 is symmetric, any permutation of a, b, c will also be a valid solution. Since N is positive, zero denominators are only of concern when $x = 4$. We will see later on that this is not a problem.

Definition 2.1. A **projective space** is a space where each point corresponds to the line passing through that point and the origin in the Euclidean space one dimension higher.

Definition 2.2. A **projective plane** is a two-dimensional projective space.

Note that points (x, y) are on the projective plane, so if (a, b, c) is a solution, then (ta, tb, tc) is also a solution for any constant t. The inverse transformation formulas above only guarantee that the values of a, b, and c are rational, not integer, so we must multiply them by the least common multiple of the denominators of the three fractions to get integer values.

Definition 2.3. The **order** of an element a of a group is the smallest positive integer m such that $m * a = e$, where e is the identity element of the group. Over the group of rational points on an elliptic curve, denoted $\mathbf{E}(\mathbf{Q})$, the identity is the point at infinity. If no finite m exists, then a is defined to have infinite order.

Definition 2.4. The **torsion subgroup** of an elliptic curve E over the rationals is the subgroup of points on E that have finite order with respect to the group law of addition over an elliptic curve.

Definition 2.5. (The Point Duplication Formula) If $P = (x, y)$, then the x-coordinate of $2P$ is

$$\frac{x^4 - 2ax^2 - 8bx + a^2}{4(x^3 + ax + b)}.$$

From this, it is easy to find the y-coordinate.

Theorem 2.6. *The torsion subgroup of this elliptic curve is isomorphic to $\mathbb{Z}/6\mathbb{Z}$ for all $N > 2$.*

Proof. We refer to [BM14].

Lemma 2.7. *When $N > 2$, the curve has exactly one point of order 2.*

Proof. We note that rational points $T_2 = (x, y)$ of order 2 must satisfy $y = 0$. In order for there to be three distinct rational points of order 2, the equation $x^3 + (4N^2 + 12N - 3)x^2 + 32(N + 3)x$ must have three rational roots. Since $(0, 0)$ is a point on the curve, the equation must be divisible by x, which we divide out to get

$$x^2 + (4N^2 + 12N - 3)x + 32(N + 3).$$

In order for a quadratic to have two rational solutions, its discriminant must be a perfect square. This quadratic has discriminant $(2N - 3)(2N + 5)^3 = (2N + 5)^2(2N + 5)(2N - 3)$, so $(2N - 3)(2N + 5) = (2N + 1)^2 - 16$ must be a perfect square for there to be more rational points of order 2. The only integer solution is $N = 2$. (When $N > 2$, the difference between perfect squares is too large for $(2N + 1)^2 - 16$ to be a square). Therefore, when $N > 2$, the curve has only one point of order 2, which is $(0, 0)$. ■

Lemma 2.8. *The curve has exactly two rational points of order 3, which are the inflection points.*

The proof of this is an exercise that the reader has likely already completed. It is given in Chapter 14 of Simon's cryptography book. [RS02]

Because we have points of order 3 and points of order 2, there must exist points $T_6 =$ (x, y) of order 6. To find them, we substitute the coordinates of T_3 into the point duplication formula to get

$$T_6 = (8(N+3), \pm 8(N+3)(2N+5)).$$

Lemma 2.9. *There exists no point of order 12.*

Proof. Plugging in the coordinates of T_6 into the point duplication formula reveals that a point T_{12} could only exist if the elliptic curve is degenerate, meaning that it has a double or triple root x_0 and a singularity at the point $(x_0, 0)$. The full proof of this lemma is given in Brenmer and MacLeod's paper [BM14] ■

Given that there exists exactly one point of order 2 and two points of order 3, and no points of order 12, we can conclude that the torsion subgroup of this elliptic curve is isomorphic to $\mathbb{Z}/6\mathbb{Z}$. ■

Although each rational point on the elliptic curve corresponds to a solution to the original Diophantine equation, torsion points either give degenerate solutions or no solutions at all (for example, when $x = 4$, it's not possible to obtain values for a, b, and c because doing so would involve dividing by zero). Therefore, we need points of infinite order for a general solution, and furthermore, the elliptic curve must be of nonzero rank. Given a generator point G, we calculate successive multiples of G until we find one that corresponds to a positive solution to Equation 1.1.

Example. Let's look at solutions to the case of $N = 6$. We plug $N = 6$ into the substitutions given in section 2, which gives

$$x = \frac{-36(a+b+2c)}{(2a+2b-c)+6(a+b)}, \; y = \frac{612(a-b)}{(2a+2b-c)+6(a+b)}$$

The elliptic curve is then $y^2 = x^3 + 213x^2 + 288x$, with inverse transformations

$$\frac{a}{a+b+c} = \frac{72-x+y}{18(4-x)}, \; \frac{b}{a+b+c} = \frac{72-x-y}{18(4-x)}, \; \frac{c}{a+b+c} = \frac{-36-8x}{9(4-x)}.$$

Since these solutions are projective, we simply multiply through by the least common multiple of the three denominators to find integer solutions when converting back.

I ran a simple Sage program to find a generator point, which returned $P = (-200, 680)$. This corresponds to

$$\frac{a}{a+b+c} = \frac{7}{27}, \; \frac{b}{a+b+c} = \frac{-1}{9}, \; \frac{c}{a+b+c} = \frac{23}{27}.$$

Multiplying by 27, we have the integer solution

$$a = 7, b = -3, c = 23.$$

(any of its cyclic permutations and/or multiples will also be solutions). However, it's not a positive solution, so it's not what we want. Trying this with $2P$, $3P$, and so on, all result in nonpositive solutions, until $11P$, which is the first positive solution, with

$$\frac{a}{a+b+c} = \frac{215414330946641634663592904941241100211134475343849362786299984417416712566876348555729759}{2522933318618339743965107094609818107320447277182370195426878680840446278374765133319886 49},$$

$$\frac{b}{a+b+c} = \frac{2174441710508871062956912130261577284476310982920530271138961051179186919439473867681552 7}{22929362522934293171051183947090378609069448369721321834263721606255860388989394946501730 7},$$

$$\frac{c}{a+b+c} = \frac{4497836684187902457419534185672218724065611288993498002492319265501461434113932335764023}{876038696131594590943130797193354012613895908020020495283044591884234265542591295465920 27}.$$

Multiplying by the least common multiple of the denominators gives the minimal solution
$a =$ 20260869859883222379931520298326390700152988332214525711323500132179943287700005601210288 7971538685332071313024772694704508282339365 57,
$b =$ 22503240220126838668864264619424948111412000849212232184619673775885644776162207677896322 5735852195244304981379971238636762392597144 7,
$c =$ 12183432427029058557922642378688032230730902983101212975267528305583238455039100718519992 1795970402428069975929055900916203510297402 3.

3. Nonexistence of solutions for odd N

Theorem 3.1. *When N is odd, the equation has no solutions.*

The proof is beyond the scope of this paper, as it is rather long and involves much casework. You can view it in Brenmer and MacLeod's paper. [BM14]

4. Sizes of minimal solutions

Brenmer and MacLeod's paper [BM14] contains a table of minimal solution sizes, for the range $4 \leq N \leq 200$ where a solution size is defined as the maximum number of digits in the numbers a, b, and c. It also shows the corresponding values of m, where m is the smallest integer such that the point $m * P$ corresponds to a positive integer solution.

Let's look at a couple of examples, which will be relevant in the next section:
When $N = 178$, the smallest positive integer solution is at $m = 2945$, with a solution size of 398605460 digits. When $N = 896$, the smallest positive integer solution is at $m = 161477$, with a solution size in the trillions of digits.

5. Computational Complexity Analysis of Searching for Solutions

We will discuss the time complexity of calculating solutions to (1.1), and determine when it is computationally feasible to do so.

Let's first start by introducing Hilbert's 10th problem, which asks whether there exists an algorithm for determining the solvability of general Diophantine equations. The problem was solved in 1970; it was proven that no such general algorithm exists. [Dav73]

Theorem 5.1. *There is no computable upper bound for the minimal solutions to (1.1)*

Proof. This is a simple corollary of Hilbert's 10th problem; we proceed using proof by contradiction. Let's assume that there is a computable upper bound. Then, there exists an algorithm for finding solutions: Test all possible positive integer values of a, b, and c below the upper bound. This is a contradiction, because Hilbert's 10th problem being proved unsolvable implies that there exists no general algorithm for solving Diophantine equations. ■

First, let's look at the complexity of operations on integers and rational numbers. For two n-digit integers, addition and subtraction can be done in $O(n)$, and multiplication and division in $O(n \log n)$. Finding the GCD can be done in $O(\log n)$ by the Euclidean algorithm. When we are working with rational numbers with n-digit numerators and denominators, all four operations are $O(n \log n)$ because the cost of these operations is dominated by multiplication.

Next, we will examine the elliptic curve operations. Let n be the number of digits in the coordinates of each point. Point addition requires only a constant number of multiplications, so its complexity is $O(n \log n)$. Point multiplication by a number k can be done in $O(\log k)$ additions by repeated doubling (similar to repeated squaring in fast exponentiation), so its complexity is $O(n \log n \log k)$.

Finally, we would like to know how many multiples of G are required to find a point corresponding to positive solutions. In order for it to correspond to a positive solution, a point must lie on one of two sections on the bounded section of the elliptic curve. Since rational points are dense on both components of the elliptic curve [Hur17], we can assume that the points $m * P$ generated by P are equally distributed about the bounded part of the elliptic curve. Moreover, using arc length estimates, the probability that any $m * P$ lies on one of these sections is $O(1/N)$. [BM14] It follows that the minimum value of m required to find the smallest positive solution is $O(N)$.

Now that we've introduced the necessary background, let's outline the steps of the algorithm and the complexity of each. We define d to be the maximum number of digits in a, b, and c.

1. Transform the equation into an elliptic curve, using the substitutions given earlier in section 2. This can be done in a constant number of arithmetic operations using the number N, so this is done in $O(N \log N)$.

2. Find a generator point on the elliptic curve. The computational complexity of this depends on its height, which can be expressed in terms of the Gross-Zagier Formula. [Han11] However, in addition to being well beyond the scope of this paper, it's also not helpful for defining the time complexity in terms we can express. Furthermore, since d is the logarithm of the uncomputable upper bound and thus grows extremely fast, we'll assume that the complexity of finding a generator is not greater than the complexity of arithmetic operations on d-digit rationals.

3. Calculate successive multiples of G until one corresponds to a positive solution to the original equation. Let m be the smallest positive integer such that $m * G$ corresponds to a valid solution. We know that m is probabilistically $O(N)$. Let d be the maximum number of digits among a, b, and c in such a solution. For each multiple of G, we perform a constant number of arithmetic operations on rationals of maximum size d, so this is done in $O(m * (d \log d)) = O(N * (d \log d))$.

Due to the rapid growth of d as N increases, the $N * (d \log d)$ term dominates all others. Hence, the time complexity of the entire algorithm is $O(N * (d \log d))$.

Summit, the fastest supercomputer in the world as of 2019, can handle $2 * 10^{15}$ operations per second. For the case of $N = 178$, d is $3.5 * 10^8$, so this algorithm requires a constant multiple of $N * d \log d = 1.8 * 10^{12}$. Due to the rapid growth of d, we see that this algorithm quickly becomes infeasible in the general case when N exceeds 200. As a result, we see that despite the existence of an algorithm for finding solutions, the size of the numbers involved renders the equation $N = \frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b}$ practically unsolvable for large values of N.

References

[BM14] Andrew Brenmer and Allan MacLeod. An unusual cubic representation problem. *Annales Mathematicae et Informaticae*, 43:29–41, 2014.

[Dav73] Martin Davis. Hilbert's tenth problem is unsolvable. *The American Mathematical Monthly*, 80(3):233–269, 1973.

[Han11] David Hansen. The Gross-Zagier formula. *Boston College*, 80, 2011.

[Hur17] A. Hurwitz. Über ternäre diophantische gleichungen dritten grades. *Vierteljahrschrift d. Naturforsch. Ges. Zürich*, 62(2):29–41, 1917.

[RS02] Simon Rubinstein-Salzedo. *Cryptography*. Springer Undergraduate Mathematics Series. Springer, 2002.

Email address: darren.yao@gmail.com