A Repulsion Motif in Diophantine Equations

Graham Everest and Thomas Ward

Abstract

Problems related to the existence of integral and rational points on cubic curves date back at least to Diophantus. A significant step in the modern theory of these equations was made by Siegel, who proved that a nonsingular plane cubic equation has only finitely many integral solutions. Examples show that simple equations can have inordinately large integral solutions in comparison to the size of their coefficients. A conjecture of Hall attempts to control this by bounding the size of integral solutions simply in terms of the coefficients of the defining equation. It turns out that a similar phenomenon seems, conjecturally, to be at work for solutions which are close to being integral in another sense. We describe this conjecture as an illustration of an underlying motif – repulsion – in the theory of Diophantine equations.

1 Challenging Questions.

In 1657, Pierre de Fermat challenged the English mathematicians Sir Kenelm Digby and John Wallis to find all the integer solutions to the equation

$$y^2 + 2 = x^3. \tag{1}$$

We can make an educated guess about his motivation.

Diophantus [32, Book VI, Prob. 19] asked for a right triangle, the sum of whose area y and hypotenuse h is a square, and whose perimeter is a cube. Taking the legs to be 2 and y, and $h + y$ to be 25, he noted that the square 25, when added to 2, gives the cube 27. The two equations $h^2 = y^2+4$ and $y+h = 25$ give $y = \frac{621}{50}$ as a solution to the problem. This is pleasing, but we are more interested in the intermediate observation, namely a solution in integers to the Diophantine equation (1).

Bachet, in his famous 1621 Latin translation of the *Arithmetica* of Diophantus, noted that from the one solution $5^2 + 2 = 3^3$ other rational numbers r may be found with the property that $r^2 + 2$ is a cube of a rational number. Fermat acquired a copy of this work, and by 1636 had studied it carefully and made significant advances. He also recorded several of his most influential marginal notes in this book, including the comment concerning Fermat's last theorem.

In 1656 the English mathematician Wallis, at the time chief cryptographer to Parliament, published his *Arithmetica infinitorum.* Digby brought this to the

attention of Fermat, prompting Fermat to begin a correspondence with "Wallis and other English mathematicians" concerning some of his number problems. This correspondence all passed through Digby's hands, and thus in 1657 Fermat challenged Digby and Wallis:

> Can one find in integers a square other than 25, which, when increased by 2, makes a cube? At first sight this appears difficult. Bachet's method gives infinitely many solutions in fractions, but the setting of the integers, undoubtedly beautiful and subtle, was not developed by Bachet nor in any other writings known to me.

That is, Fermat asked them to find all the integer solutions to the equation (1).

The solution $(x, y) = (3, 5)$ noted by Diophantus is one, and Digby and Wallis might have agreed it is the only one, since negative numbers had yet to enjoy a fully equal status with their positive siblings. Presumably the challenge was to show that, sign apart, there are no others, as Fermat claimed a proof that this solution is unique. It is not clear from this distance in time if any of the three protagonists ever did have a complete proof of this.

1.1 What if 2 had been 15?

What might have been the reaction if Fermat had instead challenged the English mathematicians with the equation

$$y^2 = x^3 + 15?$$

The solution $(1, 4)$ is easy to spot. It is less easy to notice that $(109, 1138)$ is also a solution – and not at all obvious that, issues of sign aside, there are no others.

Even more challenging, what about the equation

$$y^2 = x^3 + 24?$$

The solution $(1, 5)$ is once again easy to spot, as is the solution $(-2, 4)$, and it is not too difficult to find a third solution $(10, 32)$. However, life is surely too short to find the solution $(8158, 736844)$ without mechanical help. Once again, up to sign, this is the full list of solutions.

As these examples show, the effort involved in finding the solutions to the Diophantine equation $y^2 = x^3 + d$ depends in a rather unpredictable way on the constant d. Among the many results on this problem, some stand out. Euler showed that there are no positive rational solutions apart from $(2, 3)$ when $d = 1$ using the method of descent, and a long series of other special cases were solved by many other mathematicians (see Dickson [7, Chap. XX] for the details). A snapshot of the state of knowledge on this question in 1914 may be found in the work of Mordell [20], where many but not all values of d for which there are no integral solutions are found.

Nowadays the numerical facts above – not only the stated solutions, but the much deeper claim that there are no other integral solutions – can be checked

easily using one of the many sophisticated computational packages available, such as MAGMA [3]. This is greatly to be celebrated, although the facility with which these calculations can now be done risks obscuring the remarkable achievements made in Diophantine analysis over the last forty years that have made this possible. It is not our intention to survey these achievements (a detailed overview may be found in the lovely monograph of Hindry and Silverman [17]), although some highlights on the theoretical side will appear naturally.

1.2 Satisfying Fermat

Out of a patriotic desire to satisfy Fermat, we start with a (now elementary) proof that $(3, \pm 5)$ are indeed the only integral solutions to $y^2 = x^3 - 2$. The language of (and one result from) modern algebra makes this straightforward. Plainly x must be odd, for otherwise $y^2 \equiv 2 \pmod 4$, which is impossible. The ring $R = \mathbb{Z}[\sqrt{-2}]$ is a unique factorization domain (indeed, is a Euclidean domain – see Hardy and Wright [15, §14.7] for an account) whose only units are ± 1. Factorizing there gives

$$(y + \sqrt{-2})(y - \sqrt{-2}) = x^3.$$

In the ring R, $\gcd(y + \sqrt{-2}, y - \sqrt{-2}) \mid 2\sqrt{-2}$. Since x is odd, the greatest common divisor is a unit, and so must be ± 1. It follows that each factor is a unit multiple of a cube in R. Write (absorbing $-1 = (-1)^3$ if necessary)

$$y + \sqrt{-2} = (a + b\sqrt{-2})^3$$

with $a, b \in \mathbb{Z}$. Comparing coefficients of $\sqrt{-2}$ gives

$$1 = 3a^2b - 2b^3 = b(3a^2 - 2b^2)$$

which forces $b = 3a^2 - 2b^2 = \pm 1$. Solving gives $b = 1$ and $a = \pm 1$, and the two choices give $y = \mp 5$ and $x = 3$.

2 Siegel's Theorem

The following is a special case of a wonderful and deeply influential result, proved by Siegel [29] in 1929. Siegel's result was far more general, showing that all curves with positive genus have finitely many integral points, but for the case at hand may be stated as follows. (The genus is a measure of the complexity of a curve, and curves of genus zero may have infinitely many integral points.)

Theorem 2.1. *Given an integer $d \neq 0$, the equation*

$$y^2 = x^3 + d \tag{2}$$

has only finitely many integral solutions.

This specific result was known considerably earlier. Mordell [22] reported to the London Mathematical Society in 1918 that an earlier result of his [21], in conjunction with Thue's work [33], would show Theorem 2.1. Once again there is an important letter, sent by Siegel to Mordell in 1925 (an extract appears in [28]), outlining Siegel's ideas which eventually gave the general result that any nonsingular cubic curve has only finitely many integer points [29]. In honour of Mordell's contribution to this subject, the equation (2) in Theorem 2.1 is often known as *Mordell's equation*.

Even in its simplest form, Siegel's proof was recognized as *noneffective*, a term whose meaning will be discussed shortly. Siegel gave a second proof of finiteness using unique factorization, in a manner very similar to the proof in the case $d = -2$ above, by working in a suitably large ring and then reducing to a number of simpler equations called S-unit equations. This approach led to an *effective* proof, following Alan Baker's seminal work on transcendence theory. The terms effective and noneffective, although ubiquitous in number theory, are never precisely defined. A definition for general use might read as follows.

> effective *(adj.): adequate to accomplish a purpose; producing the intended or expected result;* 1350–1400; ME fr. L. *effectivus* = practical.

One might expect that an effective proof is one which produces an algorithm to implement the conclusion of a theorem. In our context, an effective proof might consist of a bound on the size of the largest solution. This would allow all solutions to be found, by simply checking integers below that bound to see if they satisfy the equation. This sounds desirable and practical, but there are examples where the gap between what theory provides and what is practical remains large.

Theorem 2.2 (Baker [1], 1968)**.** *Any integral solution of* (2) *satisfies*

$$\log |x| < 10^{10}|d|^{10^4}.$$

Even checking Euler's result that $x = 2$ gives the only positive integral solution to $y^2 = x^3 + 1$ is not computationally feasible using Baker's bound. So, despite the etymology, *effective* as used in number theory (which would certainly include the statement in Theorem 2.2) does not always mean practical.

Of course Theorem 2.2 was never intended to be a practical tool, and was never claimed to be one. It merely says that with the methods then available, this is the best bound obtainable, and sets a challenge for future generations. Dramatic improvements followed fairly quickly. For example, Stark [31] shows that

$$\log \max\{|x|, |y|\} \leqslant C|d|^{1+\epsilon},$$

where $C = C(\epsilon)$ is an effectively computable constant depending only on ϵ. This is a considerable improvement upon Baker's bound. However, the size of $C(\epsilon)$ is necessarily large, and checking all the values of x and y below the bound is not computationally feasible, even for small values of $|d|$.

Despite the power of these results and the direction of research initiated by the work of Baker and Stark, the modern computational facility for solving Mordell's equation came about *not* by further reducing the size of the bounds above. The method is actually less direct; see the work of Gebel, Pethö, and Zimmer [12] for complete details about the toolkit now used. In a later paper [13], the same authors show how this method resolves the equation for $|d| \leqslant 10^4$, and for almost all d with $|d| \leqslant 10^5$.

Today, after major theoretical and computational improvements, computer packages will find all the integral solutions of equations $y^2 = x^3 + d$ provided d lies within *reasonable* bounds. To give an idea of what counts as *reasonable* it is worthwhile doing some experiments yourself.

3 Hall's Conjecture

To reiterate, solving Mordell's equation in practice does not rely upon obtaining a very strong upper bound for the size of the largest solution, then checking smaller solutions on a computer. In fact an important, simple, natural, question remains unsolved to this day.

Question 1. What is the best theoretical bound (in terms of d) for the size of the largest integer x solving $y^2 = x^3 + d$ in integers?

The truth is that the best known bound is far from what seems likely to be true. What seems likely to be true is the subject of the following conjecture made by Hall [14], which we will now discuss. Hall's conjecture has been subject to extensive numerical checking, but a proof seems to require dramatically more powerful methods than those currently available.

Conjecture 1 (Hall)**.** Given $\epsilon > 0$, there is a constant $C = C(\epsilon)$ such that, for any nonzero $d \in \mathbb{Z}$, any integral solution of $y^2 = x^3 + d$ satisfies

$$\log |x| < (2 + \epsilon) \log |d| + C.$$

Originally, Hall [14] conjectured the same bound but with $\epsilon = 0$. This is no longer thought to be likely.

The audacious nature of the conjecture is not immediately apparent. Our second and third examples ($d = 15$ and $d = 24$) show that simple equations with small coefficients can have enormously large integral solutions. What Hall's conjecture suggests is that, when properly calibrated, the phenomenon of large integral solutions of an equation with small coefficients is not beyond constraint.

The conjecture of Hall follows from the infamous *ABC conjecture*, formulated by Masser and Oesterlé [25] in 1985, about the relative sizes of a zero sum of three integers. Write

$$r(N) = \prod_{p|N} p$$

(where the product is taken over the prime factors of N) for the *radical* of an integer N. The ABC conjecture says that for any $\epsilon > 0$ there is a constant $K(\epsilon)$ such that, whenever

$$A + B + C = 0$$

in nonzero coprime integers A, B, C, we have

$$\max\{|A|, |B|, |C|\} \leqslant K(\epsilon) r(ABC)^{1+\epsilon}.$$

To see how this relates to the Mordell equation, assume that x and y are integers with $y^2 = x^3 + d$. Then for any $\epsilon > 0$ the ABC conjecture implies that

$$|x|^3 \leqslant \max\{|x|^3, y^2, d\} \leqslant K(\epsilon) r(x^3 y^2 d)^{1+\epsilon} \leqslant K(\epsilon) |xyd|^{1+\epsilon}.$$

The Hall bound follows by taking logs and noting that $|y|$ is approximately $|x|^{3/2}$.

Hall's conjecture has been extensively tested. Table 1 shows values of integers x and d with an integral y satisfying $y^2 = x^3 + d$ having $\log x$ large in comparison with $2 \log |d|$. It is taken from Elkies' website [8] (see also the paper [13]). The table is surprising in two opposite senses: firstly, it gives more examples of inordinately large solutions of simple Diophantine equations, and secondly, it shows that they nonetheless fall within sight of a reasonable constraint upon how large they could be when viewed on a logarithmic scale.

Table 1: Large values of $\log x / 2 \log |d|$ in Hall's conjecture.

d	x	$\log x$	$\log x / 2 \log \lvert d \rvert$
-1641843	5853886516781223	36.305	1.268
-30032270	38115991067861271	38.179	1.108
1090	28187351	17.154	1.226
193234265	810574762403977064	41.236	1.080
17	5234	8.562	1.511
225	720114	13.487	1.245
24	8158	9.006	1.417
-307	939787	13.753	1.200
-207	367806	12.815	1.201
28024	3790689201	22.055	1.076

Does Table 1 convince? Taken together with the implication of the ABC conjecture, the answer is probably yes, in the sense that this is evidence for a sensible conjecture. We have presented it because, later on, two more tables will appear and a direct comparison will be invited.

4 Repellent Powers

It might not be apparent so far, but the point at infinity is enormously important in understanding solutions of Mordell's equation. For example, the set

of rational points on the curve forms a group under a natural geometric form of addition, which will be described in Section 5. The point at infinity is the identity element for this group operation.

From our viewpoint, Siegel's theorem may be interpreted to say that the point at infinity *repels* the integer points on the curve $y^2 = x^3 + d$. In other words, there is a (punctured) neighborhood of the point at infinity free of integer points. Empirically, we might even say we observe integer points repelling each other. For a sophisticated instance of this repulsion property being used to understand integral points on elliptic curves, see the recent work of Helfgott and Venkatesh [16]. Their methods will yield explicit constants that might well quantify practically the rate at which integral points repel each other coordinate-wise. This repulsion between integral points will be something of a mantra throughout this paper, and will inform the latter part significantly. For now though, consider the idea of points with fixed arithmetic properties repelling each other as a kind of paradigm for understanding other results in Diophantine equations.

The results stated so far may be seen as an instance of a general tendency for distinct *integral powers* to repel each other. Thus, for example, Baker's result may be phrased as follows. If x and y are positive integers with $y^2 \neq x^3$ then there is a constant $C = C(x)$ with

$$|y^2 - x^3| > C(\log x)^{10^{-4}};$$

Stark's result says that for any $\kappa < 1$ there is a constant $C = C(\kappa)$ with

$$|y^2 - x^3| > C(\log x)^{\kappa},$$

and Hall's conjecture says that for any $\epsilon < 1/2$ there is a constant $C = C(\epsilon)$ with

$$|y^2 - x^3| > C x^{\frac{1}{2}-\epsilon}.$$

These are all statements about (subsequences of) the sequence

$$a = (a_n) = (1, 4, 8, 9, 16, 25, 27, 32, 36, 49, 64, 81, 100, 121, 125, \dots)$$

of perfect powers: numbers of the form n^m with $n, m \in \mathbb{N}$ and $m \geqslant 2$. A gap of one is seen early on, and the conjecture of Catalan [6] is that 8 and 9 are the only consecutive pair in the sequence a. The proof of this result followed a path that once again illustrates the slightly ambiguous way in which the word *effective* is used. Tijdeman [34] used sharpened versions of Baker's theorem to find a number T with the property that any positive integral solution (x, y, m, n) to the Diophantine equation $x^m - y^n = 1$ must have $\max\{x, y, m, n\} \leqslant T$. This meant that a proof of Catalan's conjecture was reduced to a finite list of possibilities to check – surely the most effective of effective statements. Unfortunately a by-product of the transcendence methods used was that the number T was enormous, leaving a finite but hopelessly impractical calculation to be done. The conjecture was finally proved by Mihăilescu [19] using a mixture of analytic and algebraic methods.

Pillai studied many properties of the sequence of perfect powers, and in a paper [26] of 1945 wrote

> I take this opportunity to put in print a conjecture which I gave during the conference of the Indian Mathematical Society held at Aligarh. Arrange all the powers of integers like squares, cubes, etc. in increasing order [...]. Let a_n be the nth member of this series [...]. Then
> $$\liminf_{n\to\infty}(a_n - a_{n-1}) = \infty.$$

The conjecture of Pillai is exactly equivalent to the conjecture that for any $k \geqslant 1$ the Diophantine equation $x^m - y^n = k$ has only finitely many solutions. This remarkable problem remains open, and is the subject of a recent survey by Waldschmidt [35].

4.1 The Gap Principle

The phenomena we are describing in this paper will be familiar to workers in Diophantine equations, although possibly under another name: the gap principle. Again, this is not formulated precisely anywhere that we can find. Roughly stated, it says that where Diophantine phenomena occur (say, as rational solutions to a Diophantine equation or to an inequality) subject to some reasonable constraint, they will respond by exhibiting measurable gaps. This data is then fed back into the technicalities of the argument. Strictly speaking, there is no single gap principle; it is more of a style of argument. We invite readers to explore the current literature on Diophantine equations to see where this term – or this phenomenon – occurs.

A simple observation of this type is that the denominators of a sequence of good rational approximations to a real number must grow rapidly. To make this precise, assume that $\alpha \in \mathbb{R}$ is a real number with the property that there is an infinite sequence of rational approximations $\left(\frac{p_n}{q_n}\right)$ in lowest terms with

$$\left|\alpha - \frac{p_n}{q_n}\right| \leqslant \frac{1}{q_n^{2+\epsilon}}$$

for some fixed $\epsilon > 0$ and with $q_{n+1} > q_n$ for all $n \geqslant 1$. Then, by the triangle inequality,

$$\frac{1}{q_n q_{n+1}} \leqslant \left|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\right| \leqslant \left|\alpha - \frac{p_n}{q_n}\right| + \left|\alpha - \frac{p_{n+1}}{q_{n+1}}\right| \leqslant \frac{2}{q_n^{2+\epsilon}}.$$

It follows that $q_{n+1} \geqslant q_n^{1+\epsilon/2}$ for sufficiently large n. A consequence of this observation is that there must be extremely rapid growth in the size of q_n: there is some constant $c > 0$ for which $q_n > \mathrm{e}^{\mathrm{e}^{cn}}$.

We mention two further instance of gap principles, one old and one recent and very germane. Ingram [18] showed that large gaps occur between integral

multiples of a point P on a Mordell curve in the following sense. If $n_1 < n_2$ and $n_1 P, n_2 P$ (in the group theory sense) are both integral points, then $a^{n_1^2} < n_2$ for some explicit constant $a > 1$. This is a strong repulsion property for integer points along a sequence (nP) and, just as with the results in [16], might well translate back into a good bound for the corresponding coordinates. An instance of the kind of applications Ingram obtains is that on Mordell curves of the form (2), for all large enough d with d sixth power–free (that is, not divisible by a^6 for any integer $a \geqslant 2$), there is at most one $n \geqslant 3$ such that nP is integral (see [18, Proposition 15]).

An earlier – and highly influential – manifestation of a gap principle occurs in a paper of Mumford [24]. His result was about rational points lying on more complicated plane curves. Complicated here means not just higher degree but higher *genus*, a geometric measure of complexity – for example, Mumford's results apply to equations $y^2 = f(x)$ where $f(x) \in \mathbb{Q}[x]$ has degree at least 5, and $f(0) \neq 0$. He showed that the rational solutions exhibit naturally occurring, expanding gaps (when viewed in the most naïve sense, so that the numerators and denominators grow large very quickly). Although in one sense his result was superseded by Faltings' general proof that such curves contain only finitely many rational points, packages such as MAGMA [3] will now enumerate rational points on higher-genus curves with some ease. The repelling nature of them brings Mumford's remark vividly to life.

Example 4.1 (Taken from [4])**.** The genus-2 curve $y^2 = x^6 + 1025$ has the following rational points, together with sign changes:

$$(2, 33), \left(\frac{5}{2}, \frac{285}{2^3}\right), (8, 513), \left(\frac{1}{4}, \frac{2049}{4^3}\right), \left(\frac{20}{91}, \frac{24126045}{91^3}\right)$$

(and may have others).

5 Generalizing Siegel's Theorem and Hall's Conjecture

We will now describe a recent attempt to generalize both Siegel's theorem and Hall's conjecture in one go. We need to begin by talking a little about rational solutions of our equations. Although it is not obvious, there are infinitely many rational solutions to each of our three starting equations. In each case, there is a way to produce them all, starting with a finite set of rational points. This statement summarizes enormous theoretical and practical knowledge about rational solutions (as opposed to integral solutions) of equations defining elliptic curves. For example, the rational points on

$$y^2 = x^3 + 15. \tag{3}$$

are all generated from the two points $(1, 4)$ and $(\frac{1}{4}, \frac{31}{8})$, using the chord-and-tangent method of constructing new points. The essential observation is that

the line joining any two rational points on the curve (or the tangent to a single rational point) intersects the curve again at a third rational point. For example, the line joining our two initial points meets the curve again at $(-\frac{11}{9}, \frac{98}{27})$.

This operation, together with reflection in the x-axis, allows all rational points on the curve to be found — but this is not an easy result to prove. To help orient the reader, notice that this is an instance of *Mordell's theorem* [23], which says that the set of rational solutions of such an equation is always generated from a finite set of rational solutions in the same geometric way.

Indeed, the operation of joining two points with a line, finding the third point of intersection, and then reflecting in the x-axis (extended by continuity to allow the original points to be identical) is a binary operation giving the set of rational points on the curve the structure of an abelian group (once the point at infinity is added), and Mordell's theorem states that this group is finitely generated. It must therefore have the form $\mathbb{Z}^r \times F$ for some finite group F, and r is called the *rank* of the curve. The quantity r, like d in (2) or Δ_E in Section 7, is an important measure of the size or the complexity of the curve.

To fix notation both now and for the sequel, note that if $P = (x, y)$ is a rational solution of $y^2 = x^3 + d$ and d is integral, then x^3 and y^2 have the same denominator. Thus the denominator of y^2 must be the square of a cube, and that of x^3 the cube of a square, so we may write

$$P = \left(\frac{A_P}{B_P^2}, \frac{C_P}{B_P^3} \right),$$

with $A_P, B_P, C_P \in \mathbb{Z}$, and with both coordinates of P in lowest terms.

Siegel's theorem makes a statement about the rational solutions P of the equation $B_P = 1$, which we may interpret as a statement about the solutions (x, y) under the condition that the denominator of x (or of y) is divisible by no primes.

Question 2. What can be said about the points P such that B_P is divisible by one prime?

In other words, what can be said about the set of rational solutions P with the property that B_P is a prime power? There is a name for points of this form. If the point P has B_P composed of primes from a set S then P is called an *S-integral point*. The Siegel–Mahler theorem predicts that for a *fixed* finite set of primes S, there are only finitely many S-integral points. What we are doing is slightly different, fixing the size of the set S (typically $|S| = 1$), but not its contents. Reynolds [27] proved (under a general form of the ABC conjecture) that only finitely many perfect powers will occur among the B_P, so the first interesting case is when B_P is a prime.

Sometimes, it is provable that only finitely many points P have B_P equal to a prime. Indeed, this is true of our starting equation

$$y^2 = x^3 - 2. \tag{4}$$

To see this, note that on the curve (4), all the rational points can be generated by the single point $(3, 5)$. Since this point is the image of a rational point under

a 3-isogeny, [10, Theorem 1.3] applies to demonstrate the finiteness claim (an isogeny is a map defined by rational functions with rational coefficients from one elliptic curve to another that is a homomorphism of the group law; see [30] for a detailed discussion). Inspection suggests that, in truth, none of the rational points yield prime values B_P. However – a familiar refrain – the amount of checking of small cases remains, as yet, unfeasible.

On the other hand, searching yields a large number of rational points P on equation (3) with B_P equal to a prime. The examples below, and all that follow, were obtained using searches in Pari-GP [2]. What prevents us exhibiting many more examples is lack of space, not data. Readers interested in following this up should consult [11].

Example 5.1. The following x-coordinates of rational points P on the curve

$$y^2 = x^3 + 15$$

have B_P equal to a prime: $-\frac{11}{3^2}$, $\frac{75721}{53^2}$, $\frac{578509}{367^2}$, $-\frac{349755479}{11909^2}$, $\frac{556386829130869}{17684189^2}$, and $\frac{648924294143886280569007132 81}{259476976750177^2}$.

6 The condition that B_P is a prime

Computational evidence, as well as a heuristic argument [10, 11], suggests that when the set of rational points is generated by more than one point, as in equation (3) for example, there will be infinitely many points P with B_P equal to a prime.[1] Although the papers [10, 11] contain a great deal of numerical evidence, as well as a reasoned heuristic argument, we must admit that the conjecture has not been proved, even for one curve.

The question we now ask can be put roughly as follows.

Question 3. Where are the rational points P with B_P a prime?

It makes sense to set the question in a slightly wider context, even though our main concern lies with the case stated.

Definition 6.1. The *length* of a rational point P is the number of distinct prime divisors of B_P.

Thus the integral points are precisely the length-0 points. Length-1 points are those with B_P equal to a prime power. Our comments will apply to rational points whose length is bounded, but the question above asks simply for the location of the length-1 points. As discussed in Section 4, Siegel's theorem may be interpreted to say that the point at infinity repels length-0 points. Is it possible that the same might be true for length-1 points? In other words, do length-1 points have bounded x-coordinates?

A first attack is computational. To increase the options, a search was made on curves $y^2 = x^3+Ax+B$ with $4A^3+27B^2 \neq 0$, that is, on elliptic curves (which

[1] This holds provided a technical assumption is met, namely, that the set of points should not lie in the image of a rational isogeny.

is the proper setting for this question for reasons we will expand on later). This does yield examples of length-1 points with inordinately large x-coordinates.

Example 6.2. The curve $y^2 = x^3 - 7x + 10$ contains a rational point P with

$$B_P = 1447603299835841947353852689166657347931774207 1,$$

a prime, and with

$$x(P) = 175567.984\ldots.$$

The x-coordinate has been expressed as a real number to emphasize its rough size – this is less apparent when the number is written in rational form with a numerator and denominator.

6.1 Generalizing Siegel's Theorem

Does this mean that the obvious generalization of Siegel's theorem is false? One might think so given Example 6.2 above, and others like it – but that is to miss the flow of our argument thus far. Drawing a parallel with our earlier comments, the evidence both from examples of Siegel's theorem and from the constraints along the lines of Hall's conjecture suggests that the right parameter to measure is $\log x / \log |d|$ (for a suitable notion of the parameter d adapted to more general curves). Viewed on this logarithmic scale, we present computational evidence (see Table 2) to suggest that infinity does indeed repel length-1 points. In other words, we are suggesting that the generalization of Siegel's theorem may be credibly strengthened along the lines of Hall's conjecture.

More even than this, an elliptic curve is fundamentally a projective object (in particular, the process of adding the point at infinity as the identity element can be formalized by viewing the curve as a subset of projective space). What this means is that its true nature only becomes revealed when viewed as lying in projective space. On that basis, there are no specially favored points: although the point at infinity is chosen traditionally as the identity for the group law, the group structure makes sense with any chosen rational point as identity, with the appropriate changes. Thus we are drawn, perhaps with some trepidation, towards a belief that *all* rational points (or even all algebraic points) repel the points with length below a fixed bound. In other words, around each rational point there is a punctured neighborhood free of points of bounded length.

In order to examine this conjecture, as well as to frame it along earlier lines, we introduce a suitable notion of the distance between two points.

Definition 6.3. Given a rational point Q, we define the *logarithmic distance* from $P \neq Q$ to Q to be

$$h_Q(P) = -\log |x(Q) - x(P)|$$

when Q is finite, and

$$\log |x(P)|$$

when Q is the point at infinity.

To understand the thinking behind Definition 6.3, start with the assumption that points close to the point at infinity must be considered large, and the point at infinity (in projective space) is a point like any other point. For the sake of consistency, logarithmic distances work the same way, so that points which are close to each other have a large logarithmic distance – a general feature of such distances in Diophantine geometry. To see why this is natural, notice that a measure of the size of a rational number $\frac{a}{b}$ in lowest terms is given by the *height* $H(\frac{a}{b}) = \max\{|a|, |b|\}$. Then a rational number very close to, but not equal to, a specified number necessarily has very large height.

Conjecture 2. Let k denote a fixed positive integer, and assume that d is sixth power–free. Then the set of rational points on the curve (2) repels the rational points with length below k (the "Siegel part" of the conjecture). If Q denotes a fixed rational point, then there is a constant $C = C(k, Q)$ such that

$$h_Q(P) \leqslant C \log |d|$$

for any rational point P with length below k (the "Hall part").

Notice that if Q is not the point at infinity, then the coordinates of Q determine d, so in this case the conjecture suggest that there is a constant $C = C(k, Q)$ for which $h_Q(P) \leqslant C$ for all rational points $P \neq Q$ on the curve with length below k.

The condition on d, that it is not divisible by any sixth power, is a natural one. Without this condition, we would be free to multiply the equation through by sixth powers of integers to create more and more integral and length-1 points in an artificial way. The general hypothesis, of which this is the special case for $y^2 = x^3 + d$, is that an elliptic curve be in *minimal form*.

In the next section, computational evidence for Conjecture 2 will be presented. Notice that some evidence for Conjecture 2 already comes from the data examined in the first part of the article. The conjecture predicts that integral points are not simply repelled by the point at infinity, but also by each other – exactly what we observe. Conversely, the conjecture too sheds light upon the data. Looked at this way, we should not be surprised that, when more than a couple of integral points exist, the outliers are forced to lie a long way out.

7 Computational Evidence

This was obtained in [9] for elliptic curves (see [5, 30] for background) of the form

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with $a_1, \ldots, a_6 \in \mathbb{Z}$. The role of the parameter d, measuring the complexity of the curve, is played here by the *discriminant*, a nonzero integer

$$\Delta_E \in \mathbb{Z}[a_1, \ldots, a_6].$$

This is given by a (complicated) explicit polynomial in the variables $a_1, \ldots, a_6$. Write $h_E = \log|\Delta_E|$. Each curve E is recorded as a list $[a_1, a_2, a_3, a_4, a_6]$, and Table 2 shows what we believe is the maximal distance $\overline{h}$ from infinity of the length-1 points on that curve.

The curves are all in minimal form and all have rank 2, and the search range runs over length-1 points of the form $mP+nQ$, where P and Q denote generators of the torsion-free part of the group of rational points, and $|m|, |n| \leqslant 150$. The search range is necessarily constrained because checking that points have length 1 requires primality testing on some large integers. In every case, the largest value $\overline{h}$ occurs fairly early, strengthening our belief that it is truly the maximum – see [9] for details. Conjecture 2, suitably extended to more general elliptic curves, predicts that the ratio $\overline{h}/h_E$ will be uniformly bounded. Readers are invited to compare this table of values with the one for Hall's conjecture itself (Table 1) given earlier.

Table 2: Infinity-repelling length-1 points.

E	$\lvert\Delta_E\rvert$	$\overline{h}$	$\overline{h}/h_E$
[0,0,1,-199,1092]	11022011	12.809	0.789
[0,0,1,-27,56]	107163	11.205	0.967
[0,0,0,-28,52]	236800	13.429	1.085
[1, -1, 0, -10, 16]	10700	9.701	1.045
[1,-1,1,-42,105]	750592	8.136	0.601
[0, -1, 0, -25, 61]	154368	16.592	1.388
[1, -1, 1, -27, 75]	816128	12.363	0.908
[0, 0, 0, -7, 10]	21248	12.075	1.211
[1, -1, 0, -4, 4]	892	11.738	1.727
[0, 0, 1, -13, 18]	3275	6.511	0.804
[0, 1, 0, -5, 4]	4528	7.377	0.876
[0, 1, 1, -2, 0]	389	9.707	1.627
[1, 0, 1, -12, 14]	2068	9.819	1.286

Table 3 shows some curves with $Q = (0, 0)$, and the maximal distance $\overline{h}_Q$ from Q of the length-1 points. The remarks about search ranges and our confidence about the true nature of the maximum apply here as before; see also [9]. Note that points P close to $Q = (0, 0)$ yield large values of $h_Q(P)$ because these are logarithmic distances. The thrust of Conjecture 2 is that these values, properly scaled, are not inordinately large, but satisfy a reasonable constraint. Once again, a direct comparison is invited between this table and Table 1.

Table 3: Curves with $Q = (0, 0)$.

E	$\lvert\Delta_E\rvert$	$\overline{h}_Q$	$\overline{h}_Q/h_E$
[0, 0, 0, 150, 0]	216000000	6.436	0.335
[0, 0, 0, -90, 0]	46656000	3.756	0.212
[0,0,0,-132,0]	147197952	4.470	0.237
[0,1,0,-648,0]	17420977152	0.602	0.025
[0,0,0,34,0]	2515456	2.107	0.143
[0,0,0,-136,0]	160989184	0.279	0.014
[0,1,0,-289,0]	1546140752	5.712	0.269

8 A Theorem

Given the speculative nature of this study, it is a little surprising to find certain conditions where the Siegel part of the conjecture can be proved unconditionally, and the Hall part conditionally, at least when Q is the point at infinity. What follows is a modest result, but it does yield examples where further testing may be carried out, and it provides some support for Conjecture 2. Recall that rational points Q_1 and Q_2 are *independent* in the group law on the rational points of the curve if there is no point P with $Q_1 = aP$ and $Q_2 = bP$ for integers a, b.

Theorem 8.1 (Everest and Mahé [9])**.** *Consider the equation*

$$y^2 = x^3 - Nx$$

for some positive integer N. Assume that Q_1 and Q_2 are independent rational points with $x(Q_1) < 0$ and with $x(Q_2)$ equal to a square.

1. *("Siegel part") There is a bound upon $|x(P)|$ as P runs over length-1 points in the group generated by Q_1, Q_2.*
2. *("Hall part") Assume additionally that the ABC conjecture holds in $\mathbb{Z}$. Then there is a constant K, independent of N, for which*

$$\log |x(P)| \leqslant K \log N,$$

for all length-1 rational points P.

Example 8.2. The points $Q_1 = [-9, 9], Q_2 = [49/4, -217/8]$ on the curve

$$y^2 = x^3 - 90x$$

satisfy the hypotheses of Theorem 8.1. Computations support the belief that infinitely many length-1 points lie in the group generated by Q_1 and Q_2.

Example 8.3. The points $Q_1 = [-9, 120], Q_2 = [841, 24360]$ on the curve

$$y^2 = x^3 - 1681x$$

satisfy the hypotheses of Theorem 8.1. Note that $x(Q_2) = 29^2$. Also, in this case Q_1 and Q_2 are generators for the torsion-free part of the group of rational points. As before, it seems likely that infinitely many length-1 points lie in the group generated by Q_1 and Q_2.

Dedication. The first author dedicates this paper to the staff of Mulbarton Ward and the Weybourne Day Unit in Norwich. The second author dedicates this paper to the memory of his friend and colleague Graham Everest (1957–2010).

References

[1] A. Baker, The Diophantine equation $y^2 = ax^3 + bx^2 + cx + d$, *J. London Math. Soc.* **43** (1968) 1–9.

[2] C. Batut, K. Belabas, D. Bernardi, H. Cohen, and M. Olivier, *User's Guide to PARI-GP*, Laboratoire A2X, Université Bordeaux I, Bordeaux, 1998.

[3] W. Bosma, J. Cannon, and C. Playoust, The Magma algebra system. I. The user language, *J. Symbolic Comput.* **24** (1997) 235–265.

[4] A. Bremner and N. Tzanakis, On the equation $Y^2 = X^6 + k$, *Ann. Sci. Math. Québec* (to appear).

[5] J. W. S. Cassels, *Lectures on Elliptic Curves*, London Mathematical Society Student Texts, vol. 24, Cambridge University Press, Cambridge, 1991.

[6] E. Catalan, Note extraite d'une lettre adressée à l'editeur par Mr. E. Catalan, Répétiteur à l'école polytechnique de Paris, *J. Reine Angew. Math.* **27** (1844) 192.

[7] L. E. Dickson, *History of the Theory of Numbers. Vol. II: Diophantine Analysis*, Chelsea, New York, 1966.

[8] N. D. Elkies, List of integers x, y with $x < 10^{18}$, $0 < |x^3 - y^2| < x^{1/2}$, available at `http://www.math.harvard.edu/~elkies/hall.html`.

[9] G. Everest and V. Mahé, A generalization of Siegel's theorem and Hall's conjecture, *Experiment. Math.* **18** (2009) 1–9.

[10] G. Everest, V. Miller, and N. Stephens, Primes generated by elliptic curves, *Proc. Amer. Math. Soc.* **132** (2004) 955–963.

[11] G. Everest, P. Rogers, and T. Ward, A higher-rank Mersenne problem, in *Algorithmic Number Theory—Sydney, 2002*, Lect. Notes Comput. Sci., vol. 2369, Springer, Berlin, 2002, 95–107.

[12] J. Gebel, A. Pethö, and H. G. Zimmer, Computing integral points on Mordell's elliptic curves, *Collect. Math.* **48** (1997) 115–136.

[13] ———, On Mordell's equation, *Compositio Math.* **110** (1998) 335–367.

[14] M. Hall Jr., The Diophantine equation $x^3 - y^2 = k$, in *Computers in number theory, Proc. Sci. Res. Council Atlas Sympos. No. 2, Oxford, 1969*, Academic Press, London, 1971, 173–198.

[15] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th ed., Clarendon Press, New York, 1979.

[16] H. A. Helfgott and A. Venkatesh, Integral points on elliptic curves and 3-torsion in class groups, *J. Amer. Math. Soc.* **19** (2006) 527–550.

[17] M. Hindry and J. H. Silverman, *Diophantine Geometry: An Introduction*, Graduate Texts in Mathematics, vol. 201, Springer-Verlag, New York, 2000.

[18] P. Ingram, Multiples of integral points on elliptic curves, *J. Number Theory* **129** (2009) 182–208.

[19] P. Mihăilescu, Primary cyclotomic units and a proof of Catalan's conjecture, *J. Reine Angew. Math.* **572** (2004) 167–195.

[20] L. J. Mordell, The diophantine equation $y^2 - k = x^3$, *Proc. London Math. Soc.* **13** (1914) 60–80.

[21] ———, Indeterminate equations of the third and fourth degrees, *Quart. J. of Pure and Applied Math.* **45** (1914) 170–186.

[22] ———, A statement by Fermat, *Proc. London Math. Soc.* **18** (1920) v.

[23] ———, On the rational solutions of the indeterminate equations of the third and fourth degrees, *Proc. Cambridge Philos. Soc.* **21** (1922) 179.

[24] D. Mumford, A remark on Mordell's conjecture, *Amer. J. Math.* **87** (1965) 1007–1016.

[25] J. Oesterlé, Nouvelles approches du "théorème" de Fermat, Séminaire Bourbaki, vol. 1987/88, Exp. No. 694, *Astérisque* **161-162** (1988) 165–186.

[26] S. S. Pillai, On the equation $2^x - 3^y = 2^X + 3^Y$, *Bull. Calcutta Math. Soc.* **37** (1945) 15–20.

[27] J. Reynolds, Extending Siegel's theory for elliptic curves, Ph.D. dissertation, University of East Anglia, Norwich, 2008.

[28] C. L. Siegel, The integer solutions of the equation $y^2 = ax^n + bx^{n-1} + \cdots + k$, *J. Lond. Math. Soc.* **1** (1926) 66–68.

[29] ———, Über einige Anwendungen diophantische Approximationen, in *Collected Works*, Springer–Verlag, Berlin, 1966, 209–266.

[30] J. H. Silverman, *The Arithmetic of Elliptic Curves*, 2nd ed., Graduate Texts in Mathematics, vol. 106, Springer, Dordrecht, 2009.

[31] H. M. Stark, Effective estimates of solutions of some Diophantine equations, *Acta Arith.* **24** (1973) 251–259.

[32] P. L. Tannery, ed., *Diophanti Alexandrini Opera Omnia: Cum Graecis Commentariis*, Teubner, Leipzig, 1893–1895.

[33] A. Thue, Über Annäherungswerte Alegbraischer Zahlen, *J. Reine Angew. Math.* **135** (1909) 284–305.

[34] R. Tijdeman, Some applications of Baker's sharpened bounds to Diophantine equations, in *Séminaire Delange-Pisot-Poitou 16e année: 1974/75,* Théorie des nombres, Fasc. 2, Exp. No. 24, Secrétariat Mathématique, Paris, 1975, 7.

[35] M. Waldschmidt, Perfect powers: Pillai's works and their developments (2009), available at `http://arxiv.org/abs/0908.4031`.

GRAHAM EVEREST received a Ph.D. from King's College London in 1983 and joined the University of East Anglia mathematics department the same year. He worked in number theory, specializing in the relationship between Diophantine problems and elliptic curves. He was ordained as a minister in the Church of England in 2005, and passed away in 2010.

TOM WARD received a Ph.D. from the University of Warwick in 1989. After research positions at the University of Maryland College Park and Ohio State University, he joined the University of East Anglia mathematics department in 1992. He has been Pro-Vice Chancellor (Academic) since 2008. He works, when diary permits, in ergodic theory and its interactions with number theory.
School of Mathematics, University of East Anglia, Norwich NR4 7TJ, U.K.
t.ward@uea.ac.uk