THE CONJECTURE OF BIRCH AND SWINNERTON-DYER

Günter Harder and Don Zagier

In the year 2000, Landon Clay, an American patron of mathematics, offered \$1.000.000 each for the solution of seven of the most famous open problems in mathematics. One of these is the so-called Birch–Swinnerton-Dyer Conjecture. It involves concepts which are not widely familiar and therefore may seem at first to be one of the least attractive of the seven, but in fact it is particularly beautiful and is not at all impossible for non-specialists to understand. In this article we will try to explain what this conjecture says and why it is interesting.

Diophantine equations. Next to euclidean geometry, the field of *diophantine equations* is perhaps the oldest subject studied in Western—or, for that matter, Eastern—mathematics. It is named after the Greek mathematician Diophantus, who lived in Alexandria in the 3rd century A.D. His great work, the *Arithmetika*, was lost for centuries when the library of Alexandria was destroyed, but six of its thirteen books were rediscovered in the 16th century and, in the hands of Fermat (1601–1665) and later mathematicians, sparked a period of intense study in Europe. Four more books were discovered in 1971; the other three are still missing.

A diophantine equation is an algebraic equation in unknown variables x, y, ... for which we wish to find rational or integer solutions. (Recall that integers are numbers like 26, -5 or 0, while rational numbers are fractions like $\frac{1}{2}$ or $-\frac{7}{15}$. In school we learn about integers first, but it turns out that the study of diophantine equations becomes *harder*, not easier, if we study only their integral solutions.) Many diophantine problems are classical and have played a role in the history of mathematics. Here are a few examples.

- **(i)** Find integral solutions of the equation $a^2 + b^2 = c^2$.
- **(ii)** Determine whether a given prime number can be written as a sum of two integral squares. (For instance $29 = 25 + 4 = 5^2 + 2^2$ has such a representation, while 23 does not.)
- **(iii)** Determine whether a given prime number can be written as a sum of two rational cubes. (For instance $13 = \frac{343}{27} + \frac{8}{27} = \left(\frac{7}{3}\right)^3 + \left(\frac{2}{3}\right)^3$ can be represented, while 5 cannot.)
- **(iv)** Decide whether there is a right triangle with integral side-lengths a, b and c (where c is the hypotenuse) such that both $a + b$ and c are perfect squares.
- **(v)** Determine whether a given number n can be represented as the area of a right triangle with rational side-lengths.
- **(vi)** Show that the equation $a^n + b^n = c^n$ has no solutions in positive integers if $n > 2$.

Problem **(i)**, which was studied already in Babylonian times, is motivated by "Pythagoras's" theorem (actually much older!), which says that the equation $a^2 + b^2 = c^2$ describes the relation between the sides a, b and c of a right triangle. Thus having integer solutions like $3^2 + 4^2 = 9 + 16 = 25 = 5^2$ or $8^2 + 15^2 = 64 + 225 = 289 = 17^2$ permits us to find right triangles with integral lengths (here $(3, 4, 5)$ and $(8, 15, 17)$), and by measuring them off with a rope marked by knots at regular intervals we obtain a practical way to construct a right triangle.

Problem **(ii)** was studied and solved by Fermat (and almost certainly already by Diophantus), who found that a prime $p > 2$ is a sum of two squares if and only it leaves the remainder 1

when divided by 4. Thus we can determine immediately that the prime $p = 7485704209 = 4 \times 1871426052 + 1$ has such a decomposition, even if we do not know what it is. (The answer is in fact $71425^2 + 48828^2$.) This theorem of Fermat led in the hands of Gauss, Dedekind and others to the whole modern field of algebraic number theory.

Problems **(iii)**, **(iv)** and **(iv)** are examples of so-called *elliptic curves*, which are the subject of the Birch–Swinnerton-Dyer conjecture and of this article. Problem **(iii)** was posed and studied by Sylvester in the mid-19th century. Here even determining the solution for a single prime is quite hard: the solution given above for $p = 13$ can be found by hand, but the simplest solution of $x^3 + y^3 = p$ for the "revolutionary" prime $p = 1789$ is given by

$$(x,\, y) \;=\; \left(\frac{38119538057820221}{2828707454055574}\,,\; -\frac{24606633997841365}{2828707454055574}\right)$$

and cannot be found even by computer without using a deep theory. Problem **(iv)** was posed and solved by Fermat in a letter to the priest Mersenne in 1637. Amazingly, he was able not only to find the integral solution

$$a = 1061652293520, \quad b = 4565486027761, \quad c = 4687298610289,$$

but to prove that it is the smallest! Problem **(v)** is even older, having been formulated and solved for $n = 5$ in an Arab manuscript of 972 A.D. and again in the book *liber quadratorum* by Leonardo Fibonacci (1225), where it was shown that the answer is "yes" in this case, a right triangle with area 5 being the one with side-lengths 3/2, 20/3 and 41/6.

Finally, problem **(vi)**, the most famous diophantine problem by far, was claimed to have been proved by Fermat in 1637 in an unpublished marginal comment in his copy of Diophantus's *Arithmetika*, became famous under the name "Fermat's Last Theorem", and was finally proved by Andrew Wiles in 1997.

Easy equations, hard equations, and insoluble equations. Why are the degrees of difficulty of the six problems above so different? It turns out that all diophantine equations can be divided into three classes, called "rational," "elliptic," and "of general type." Roughly speaking, these correspond to equations of degree 2, 3, and ≥ 4, respectively (although in fact some exceptional degenerate equations of degree 3 may be of rational type and some of degree ≥ 4 can be rational or elliptic). The problem of finding rational solutions turns out to be *easy* for the first class, *hard but attackable* for the second, and *in general impossible* for the third. More precisely, rational equations usually have infinitely many solutions and there is then an easy formula which gives them all. Elliptic equations can have finitely or infinitely many solutions, but there is a beautiful theory describing the structure of the solutions and a way to predict whether there are finitely or infinitely many, given precisely by the conjecture of Birch and Swinnerton-Dyer. Equations of general type have at most finitely many solutions, as was conjectured by Mordell in 1917 and proved by Faltings in 1984, and there is in general no algorithmic way to find them.

To explain this division, consider first the Pythagorean equation $a^2 + b^2 = c^2$ (problem **(i)**). If we divide by c^2, we obtain the equation $x^2 + y^2 = 1$, where $x = a/c$ and $y = b/c$ are now rational numbers. From our schooldays we know that this equation describes the set of points (x, y) lying on the unit circle with center $(0, 0)$. So we have the equivalent problem: find the points with rational coordinates (the "rational points") on the circle. This translation between algebraic and geometric problems, unknown to Diophantus, was discovered by Descartes and eventually developed into the field now called "algebraic geometry."

We can use this example to explain why the rational curves, given by second degree equations, are easy. Consider a fixed line, say the line $x = 1$, and join an arbitrary point $P = (1, a)$ on this line to the point $(-1, 0)$ on the circle (Fig. 1). It is a computation on the level of high

school algebra that the point Q where the line joining them meets the circle has coordinates $\left(\frac{4-a^2}{4+a^2}, \frac{4a}{4+a^2}\right)$. Thus when a is rational Q has rational coordinates, and we get all rational points this way. The same method, due in essence to Diophantus, works for every rational curve.

For elliptic curves, given by equations of degree 3, the method fails but a modification of it, discovered in special cases by Diophantus and developed further by Fermat ("method of infinite descent") and his followers, can be applied: if P and Q are two known solutions, then the line through P and Q meets the curve in a third point (because the degree of the equation is 3), and this point has rational coordinates if P and Q do (Fig. 2a). As a small variant, we can let the points P and Q coincide and consider the tangent line to the curve through P; again it intersects the curve in one further point which is rational if P is (Fig. 2b). It is by these methods that Fibonacci, Fermat and their followers could find the complicated explicit solutions given above.

One also sees why this method fails if the degree is 4 or bigger: if we start with two known rational points, then the line through them has at least two further intersection points with the curve, which then have no reason to be rational, while if we start with three known rational points then they usually will not lie on a line at all. So the proliferation of solutions which we found for rational and elliptic curves does not take place here.

The magic number. We have seen that the curves which have the most interesting structure from the diophantine point of view are the elliptic ones. Can one find an effective criterion, analogous to Fermat's answer to the problem **(ii)** above, which allows us to decide whether a given elliptic diophantine problem has a solution?

The Birch–Swinnerton-Dyer (BSD) conjecture in its simplest form gives us such a criterion, although it is not yet proved in all cases. We will explain this here, and then describe a more precise form in the next section.

To state their criterion, Birch and Swinnerton-Dyer had to assume that the elliptic curve being studied had a certain property called "modularity" which says that the curve can be parametrized by functions of a special sort called "modular functions" which are among the most intensively studied objects of modern number theory. Their full definition is too technical to be given here; let us only say that they have an infinite number of periodicities analogous to the familiar periodicity of the sine curve, but of a much more complicated type. By "parametrization" we mean that there exist two modular functions $x(t)$ and $y(t)$ such that as t (the "time") runs through the real numbers, the point $(x(t), y(t))$ traces out the elliptic curve. (Think of the parametrization of the circle $x^2 + y^2 = 1$ by the functions $x(t) = \cos t$ and $y(t) = \sin t$.) The existence of a modular parametrization can be checked algorithmically for any given curve, and in practise always holds, so that the restriction to modular elliptic curves was not a major problem even before, but in fact it is now known that *every* elliptic curve with rational coefficients is modular. This theorem, which had been conjectured for many years, was proved by Wiles in 1994 for a wide subclass of curves as the key step in his proof of Fermat's last theorem, and extended by Breuil, Conrad, Diamond and Taylor to the general case in 2000.

Now using the modularity of an elliptic curve E, one can associate to E a certain **magic number** which is an integer that can be calculated algorithmically. To do this, we first associate to E a certain function $L(s)$ called its *L-function*. This function is initially defined for large s by an infinite series of the form

$$L(s) \;=\; \frac{a(1)}{1^s} + \frac{a(2)}{2^s} + \frac{a(3)}{3^s} + \cdots \qquad (s > 3/2)$$

where the $a(n)$ are certain integers associated to E in an explicit way. (If n is prime then $a(n)$ is related to the number of "solutions in integers modulo n" of the equation defining the curve, and there is a simple rule to obtain all the $a(n)$ from the ones with n prime.) As an example,

for the curve $x^3 + y^3 = 5$, the L-series begins

$$L(s) \;=\; \frac{1}{1^s} \,-\, \frac{2}{4^s} \,-\, \frac{4}{7^s} \,+\, \frac{5}{13^s} \,+\, \frac{4}{16^s} \,+\, \frac{8}{19^s} \,+\, \cdots\,.$$

The L-series no longer converges if $s < 3/2$, but the modularity of E permits one to give an alternative representation of the function $L(s)$ which makes sense for all values of s and in particular at $s = 1$. This definition, though too complicated to explain here, is quite explicit and is now part of several standard software packages. The modular theory then further tells us that the value of $L(s)$ at $s = 1$ is given by the formula $L(1) = S \cdot \Omega$, where Ω is a strictly positive real number that can be calculated easily (it is the integral of a simple algebraic function) and S is an integer. For instance, for the curve $x^3 + y^3 = 5$ the computer gives $L(1) = 1,033136608569$ and $\Omega = 1,033136608569$, so $S = 1$, while for $x^3 + y^3 = 13$ we have $L(1) = 0,000000000000$ and $\Omega = 0,751334448265$, so $S = 0$. Note that, even though $L(1)$ and Ω are known only approximately, we obtain S exactly because it is an integer. This S is the magic number.

The conjecture of Birch and Swinnerton-Dyer in its simplest form now says:

An elliptic curve has infinitely many rational points if and only if S vanishes.

We want to emphasize how truly amazing this answer is: to determine the solvability or not of a simple equation which can be explained to a high school student and of which certain cases were already known in antiquity, one applies to the equation an extremely sophisticated theory called the theory of modular functions and uses it to construct an auxiliary number whose definition is very indirect and seemingly unrelated to the equation itself, and then the properties of this number determine whether there are rational solutions. This is in itself amazing, and becomes even more so if one knows that the conjecture was formulated in the mid-1960's on the basis of beautiful but actually quite meagre experimental evidence obtained with the very primitive computers of the day. It was in fact one of the first serious conjectures in mathematics whose discovery arose from the possibility of making numerical computations electronically.

The correctness of the conjecture above is known in one direction: if S is *not* equal to zero, then the equation of the elliptic curve has only finitely many rational solutions. This was proved in 1975 by Coates and Wiles for an important subclass of curves including the problems **(iii)** and **(v)** of the introduction, and by Kolyvagin in 1988 in general.

The magic number S can not only be calculated numerically for each specific curve, but also theoretically for certain infinite families. For instance, for the problem **(v)** for n prime Tunnell showed in 1983 that $S = (N_1 - N_2)^2$, where N_1 and N_2 denote the number of ways to write n as $a^2 + 2b^2 + 8c^2$ with c even and c odd, respectively. So if $N_1 \neq N_2$, which can be checked very easily, then the answer to problem **(v)** is negative, while if $N_1 = N_2$ we expect the answer to be positive. Similarly, for Sylvester's problem **(iii)** an explicit formula for S was given by Villegas and Zagier in 1997, so here too we obtain, at least conjecturally, the complete answer to a famous classical problem.

Infinite and more infinite, zero and more zero. We have seen how to predict whether a given elliptic equation has infinitely many solutions or not. But some infinities are bigger than others. Consider, for example, Sylvester's equation $p = x^3 + y^3$ (problem **(iii)**) for the primes $p = 5$, 13 and 19. For $p = 5$ a computer search over a large range reveals no solutions, and in fact one can prove that there are none. For $p = 13$ and $p = 19$ there are infinitely many solutions but the behavior is very different: for $p = 13$ the only solutions with denominator $< 10^{14}$ are

$$13 \;=\; \left(\tfrac{7}{3}\right)^3 + \left(\tfrac{2}{3}\right)^3 \;=\; \left(\tfrac{2513}{1005}\right)^3 + \left(-\tfrac{1388}{1005}\right)^3 \;=\; \left(\tfrac{46066103}{15177582}\right)^3 \,+\, \left(-\tfrac{37397807}{15177582}\right)^3,$$

but for $p = 19$ we find already 9 solutions with denominator < 250:

$$\begin{aligned} 19 &= \left(3\right)^3 + \left(-2\right)^3 = \left(\tfrac{5}{2}\right)^3 + \left(\tfrac{3}{2}\right)^3 = \left(\tfrac{8}{3}\right)^3 + \left(\tfrac{1}{3}\right)^3 = \left(\tfrac{36}{13}\right)^3 + \left(-\tfrac{17}{13}\right)^3 = \left(\tfrac{109}{31}\right)^3 + \left(-\tfrac{90}{31}\right)^3 \\ &= \left(\tfrac{92}{35}\right)^3 + \left(\tfrac{33}{35}\right)^3 = \left(\tfrac{613}{103}\right)^3 + \left(-\tfrac{594}{103}\right)^3 = \left(\tfrac{895}{196}\right)^3 + \left(-\tfrac{831}{196}\right)^3 = \left(\tfrac{2395}{201}\right)^3 + \left(-\tfrac{2386}{201}\right)^3, \end{aligned}$$

and many more if one searches further.

The explanation of this behavior is that the number of rational points on any given elliptic curve having numerators and denominators of at most N digits grows precisely like a constant times $(\sqrt{N})^r$ as N tends to infinity, where $r \geq 0$ is a certain integer called the **rank** of the curve. Thus if the rank r is 0, there are only finitely many solutions; if r is 1, then there are infinitely many but they grow very quickly, with the nth one having numerators and denominators with roughly n^2 digits, while for larger values of r there are again infinitely many but now growing much more slowly. The ranks of the Sylvester curve $x^3 + y^3 = p$ for $p = 5$, 13 and 19 are 0, 1 and 2, respectively, and this explains the different behaviors which we just saw.

The meaning of the rank is as follows. We explained before that one can produce new points on an elliptic curve from old ones by intersecting the tangent line through a known point, or the line joining two known points, with the curve. Mordell proved in 1922 that *all* rational points on an arbitrary elliptic curve can be obtained starting with only finitely many basic solutions by repeatedly applying these two procedures (together with a symmetry, here $(x, y) \mapsto (y, x)$). The rank is essentially the minimum number of basic solutions which are needed[1]. For instance, for the curve $x^3 + y^3 = 13$ every rational point can be obtained from the single basic solution $(\frac{7}{3}, \frac{2}{3})$ by iterating the tangent and secant methods and reflecting, so the curve has rank 1. For the curve $x^3 + y^3 = 19$, all points can be obtained from the first two points $A = (3, -2)$, $B = (\frac{5}{2}, \frac{3}{2})$ on the above list by applying the same procedures ((for instance, the third point of intersection of the line $7x + y = 19$ through A and B with the curve is the point $(\frac{8}{3}, \frac{1}{3})$, the third on the list), so here the rank is 2 (Fig. 3).

On the other hand, the "magic number" S was not merely a number, but the value of a function $L(s)$ at $s = 1$ (divided by Ω), and some functions are "more zero" than others. The conjecture of Birch and Swinnerton-Dyer in its full form says that the degree of vanishing of $L(s)$ at $s = 1$ exactly determines the rank in all cases:

Conjecture of Birch and Swinnerton-Dyer (refined version). *The rank of an elliptic curve is equal to the order to which the associated L-function $L(s)$ vanishes at $s = 1$. In particular, the rank is 0 (so there are only finitely many rational solutions) if the graph of $L(s)$ does not meet the s-axis at $s = 1$; the rank is 1 if the graph meets the s-axis at $s = 1$ with non-zero slope; and the rank is ≥ 2 if the graph is tangent to the s-axis at $s = 1$.*

The three situations are illustrated in Fig. 4 for the L-functions of the Sylvester curves $x^3 + y^3 = p$ for $p = 5$, 13 and 19 ($r = 0$, 1 and 2).

The conjecture of Birch and Swinnerton-Dyer is still open, and the million dollars have not been claimed. All that one knows in general is that $r = 0$ if $L(1) \neq 0$ (as explained in the last section) and that $r = 1$ if $L(s)$ vanishes simply at $s = 1$ (Gross-Zagier (1983) and Kolyvagin (1988)). The solution of the problem, one of the deepest and most beautiful in all of number theory, will constitute a huge step forward in our understanding of the mysteries of numbers.

[1]More precisely, as was realized by Poincaré, the secant-and-tangent method leads to an an "addition" ("abelian group structure") of points on the curve which obeys the same axioms as ordinary addition: we simply declare the sum of three collinear points on the curve to be 0 (thus $P + Q + R = 0$ for the three points in Fig. 2a and $2P + R = 0$ for the three points in Fig. 2b) and define the negative of a point as its image under the symmetry. Then Mordell's theorem in mathematical language says: "the group of rational points is finitely generated, of rank r".