RENDICONTI
del
SEMINARIO MATEMATICO
della
UNIVERSITÀ DI PADOVA

J. W. S. CASSELS

Computer-aided serendipity

Rendiconti del Seminario Matematico della Università di Padova, tome 93 (1995), p. 187-197

<http://www.numdam.org/item?id=RSMUP_1995__93__187_0>

REND. SEM. MAT. UNIV. PADOVA, Vol. 93 (1995)

Computer-aided Serendipity [1] [2].

J. W. S. CASSELS (*)

Number Theory is an experimental science. Unexpected regularities are noticed in the results of calculations and only subsequently are explanations and proofs found. For example, the Law of Quadratic Reciprocity was known by experiment to Euler and others long before the first complete proof was found by Gauss: and Gauss' first proof is a verification of a kind which would not have suggested itself if the Law was not already known empirically. The advent of electronic computers has revolutionized the possibilities of numerical experimentation and so greatly increased the number of theorems suggested in this way and of conjectures still awaiting proof. Another boon of computers is that one can often easily test a conjecture suggested in other ways: it would be foolish to expend intellectual energy in attempting to prove it if a little experimentation produces a counter-example.

I am not a computer expert, but I am fortunate in having friends who are, and in having been a witness of early developments. The first approaches to electronic computing were made during the war with specialized devices for ballistic or cryptographic purposes. They lacked versatility, and if you wanted to do another calculation you had to un-

(1) «Coined by Horace Walpole upon the title of the fairy tale "The three princes of Serendip", the heroes of which "were always making discoveries, by accident and sagacity, of things they were not in quest of".» Serendip is an old name for Sri Lanka = Ceylon. An often quoted example is looking for a needle in a haystack and finding the farmer's daughter.

(2) Lecture given in Padova in October 1993 and elsewhere.

(*) Indirizzo dell'A.: Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, U.K.

plug and replug wires or something of the sort. At least in Cambridge, it is claimed that the true mark of an electronic computer is that it will read in, store and act on a program. I was present at a demonstration in, I think, May 1949 of almost the first program on the very first computer EDSAC. This was a roomful of equipment but was laughably weak by today's standards. Slowly it printed out the prime numbers up to 100. At least this is what it was supposed to do, but it printed out the squares of primes as well: the program contained a bug!

There was no such profession as computer science in those days. Several of my fellow research students in number theory found the challenges of the new discipline congenial and moved into it. I might add that I have always encouraged my research pupils to learn about computers. Apart from the more obvious advantages, if they turn out not to be creative mathematicians they will, at least, have acquired a saleable skill.

There were computers before there were electronic computers. In particular, there were machines which operated cards prepared by punching holes. The machine sorted them by sensing the holes mechanically or, later, optically. These machines were primarily for business, but were occasionally used for mathematics. One application was to Waring's Problem, that is for given n (say $n = 5$) to finding an s such that every positive integer x is the sum of at most s n-th powers of positive integers. The analytic approach naturally produces an s such that this is true for all large enough x, say for $x > x_0$. Then to get an elegant formulation one has to check for the remaining $x \leqslant x_0$. This was a drudgery for which the punched card machines were well suited. The Soviet mathematician Buchstab (1940) made a more sophisticated application. For a sieving problem he defined a function recursively: to evaluate it he used punched cards. Incidentally, the shape of the punched cards is still reflected in the procrustean format of the FORTRAN statement.

Although electronic computers were developed for less serious purposes, they were soon used for number theory. Perhaps the earliest application was by von Neumann and Goldstine (1953) to a conjecture of Kummer about the distribution of certain trigonometric sums involving a cubic residue character («Kummer sums»). Kummer had made the conjecture on the basis of hand calculations. Von Neumann and Goldstine carried the calculations much further and destroyed its plausibility. Programming was a much more exacting art then, both because of the feeble power of the machines and because the modern programming languages did not exist, so it all had to be done in «machine code». The programmer was Mrs Atle Selberg. The problem has now been cleared up by R. Heath-Brown and S. J. Patterson (1979).

Before proceeding further, we must recall some facts about elliptic curves over the rational field $\mathbb{Q}$ ([3]).

For present purposes, an elliptic curve $\mathcal{C}$ over $\mathbb{Q}$ is given by

$$Y^2 = X^3 + a_2X^2 + a_1X + a_0 , \tag{1}$$

where the $a_j \in \mathbb{Q}$ and the polynomial on the right hand side has no repeated factors. A point $\mathfrak{X} = (x, y)$ on $\mathcal{C}$ is said to be rational if $x, y \in \mathbb{Q}$. There is a single point on $\mathcal{C}$ at infinity, which is also rational by definition. A fundamental fact, which we do not prove here, is that the set $\mathfrak{G}$ of rational points on $\mathcal{C}$ form an abelian group with $\mathfrak{o}$ as zero. One puts

$$\mathfrak{x}_1 + \mathfrak{x}_2 + \mathfrak{x}_3 = \mathfrak{o} \tag{2}$$

precisely when the $\mathfrak{x}_j$ are collinear (Figure 1): in particular,

$$-(x, y) = (x, -y). \tag{3}$$

The difficult thing is to prove that addition is associative.

A celebrated result is Mordell's finite basis theorem: the group $\mathfrak{G}$ is finitely generated. This, I note in passing, is a splendid example of serendipity, though not computer-aided. He was aiming at some-

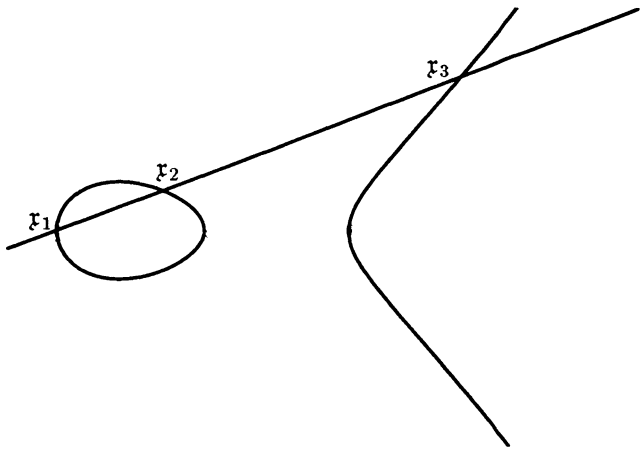

Fig. 1.

([3]) For a fuller discussion in the same vein see, for example, the author's book Cassels (1991).

thing quite different. Later, that was proved by Siegel: the finite basis theorem being a key ingredient (4).

We shall need to know a little of the ideas which enter into a modern proof of Mordell's theorem, and consider the special case

$$Y^2 = X^3 + a_2X^2 + a_1X + a_0 = (X - e_1)(X - e_2)(X - e_3), \tag{4}$$

with distinct $e_j \in \mathbb{Z}$. If (x, y) is a rational point on $\mathcal{C}$, it is easy to see that there are $u, v, t \in \mathbb{Z}$ such that $x = u/t^2$, $y = v/t^3$ are fractions in their lowest terms. It follows that

$$v^2 = (u - e_1t^2)(u - e_2t^2)(y - e_3t^2). \tag{5}$$

It is readily verified that the greatest common divisor of the first two factors on the right hand side divides $e_1 - e_2$ etc., and so, taking the squarefree kernels of the three factors we have

$$\begin{cases} u - e_1t^2 = m_1s_1^2, \\ u - e_2t^3 = m_2s_2^2, \\ u - e_3t^2 = m_3s_3^2, \end{cases} \tag{6}$$

for some $s_1, s_2, s_3 \in \mathbb{Z}$, where the triplet $m_1, m_2, m_3 \in \mathbb{Z}$ is from a finite set. On eliminating u from (6) we get the following three simultaneous quadratic equations in s_1, s_2, s_3, t, of which only two are independent:

$$\begin{cases} m_1s_1^2 - m_2s_2^2 = (e_2 - e_1)t^2 \\ m_2s_2^2 - m_3s_3^2 = (e_3 - e_2)t^2 \\ m_3s_3^2 - m_1s_1^2 = (e_1 - e_3)t^2 \end{cases} \tag{7}$$

The triplet $\{m_1, m_2, m_3\}$ corresponds precisely to the the class of $\mathfrak{x} = (x, y) \in \mathfrak{G}$ modulo $2\mathfrak{G}$. To see why there should be this link between the expressions (6) and the group structure of $\mathfrak{G}$, we suppose that $\mathcal{C}$ is given by (4) and that the points $\mathfrak{x}_j = (x_j, y_j)$ of figure 1 lie on the line $Y = lX + n$. The values of the x_j are obtained by eliminating Y, and so are given by

$$(X - e_1)(X - e_2)(X - e_3) - (lX + n)^2 = (X - x_1)(X - x_2)(X - x_3). \tag{8}$$

(4) For a historical discussion, see Cassels (1986).

On substituting an e_j for X we see that

$$(x_1 - e_j)(x_2 - e_j)(x_3 - e_j) = (le_j + n)^2 \tag{9}$$

is a square. It is not difficult to supply the rest of the argument. The proof of finite generation now uses a different type of argument that we do not need to go into here. Since it easy to determine the torsion part[5] of $\mathfrak{G}$, a knowledge of the number of triplets $\{m_1, m_2, m_3\}$ for which (7) is soluble gives the rank of $\mathfrak{G}$ (the number of generators of infinite order). Similar results apply to a general $\mathcal{C}$ not of the special form (4).

Note that we have shown that the rank of $\mathfrak{G}$ is finite, but we have no way to determine that rank, even in principle. All we have is an upper bound. To get the precise rank, we should have to decide for given $\{m_1, m_2, m_3\}$ whether there is a solution $\{s_1, s_2, s_3, t\}$ of (7) or not. For a long time there was no infallible decision procedure: as the logicians would say, the rank was not effectively computible. It is only in the last few years that this situation has begun to change.

If we can prove that some more of the equations (7) are insoluble, we get a sharper bound for the rank. We recall the theory of a single equation

$$F(X_1, \ldots, X_n) = 0\,, \tag{10}$$

where F is a quadratic form with rational coefficients. The equation

$$X^2 + Y^2 + Z^2 = 0 \tag{11}$$

has no solution[6] in $\mathbb{Q}$ as there are none in the real field $\mathbb{R}$. Similarly, the equation

$$X^2 - 2Y^2 + 3Z^2 = 0 \tag{12}$$

has no solution in $\mathbb{Q}$ on considering the behaviour[7] of a possible solution modulo powers of 3: that is, there is no solution in the 3-adic field $\mathbb{Q}_3$. We say that a Diophantine equation is «everywhere locally soluble» if the are solutions in $\mathbb{R}$ and in every p-adic field $\mathbb{Q}_p$ and that it is «soluble globally» if there is a solution in $\mathbb{Q}$. Hasse succinctly reformulated

[5] Only the 2-primary part is relevant here.

[6] As always, we exclude the trivial solution in which all the variables are 0.

[7] For if $\{x, y, z\}$ is a solution, we may suppose that $x, y, z \in \mathbb{Z}$ have no common factor. Then x, y must be divisible by 3, and (12) now implies that z is divisible by 3. Contradiction!

existing criteria for the solubility of a single equation (10), where F is a quadratic form: «If there is a solution everywhere locally, then there is a solution globally». There are other situations to which such a statement applies. It is known as a «Hasse principle» or «local-global principle».

There is no local-global principle for simultaneous equations of the type (7). For any set $\{m_1, m_2, m_3\}$ we can, however, always decide whether (7) is everywhere locally soluble. This gives a better upper bound for the rank of $\mathfrak{G}$. We can check whether this bound is the actual rank by searching (preferably by machine!) for solutions. If one can produce a solution for all the everywhere locally soluble (7), then the upper bound is the actual rank. If a thorough search fails to find the required solutions, then there is a strong suspicion that the rank is smaller than the upper bound. As a matter of experimental observation, the upper bound is usually attained. But not always. After extensive computations Selmer (1954) made the totally unexpected conjecture: *the difference between the upper bound and the actual rank is always even.*

Actually Selmer worked not with general curves (1) but with

$$X^3 + Y^3 + AZ^3 = 0, \tag{13}$$

where $A \in \mathbb{Z}$ is given. For this he found upper bounds for the rank by, methods in which, roughly speaking, multiplication by 3 [more precisely, multiplication by $\sqrt{-3}$] replaces multiplication by 2. He made extensive calculations by hand and completed them in one of the earliest uses of a computer([8]) The curve (13) is birationally equivalent over $\mathbb{Q}$ to

$$Y^2 = X^3 - 2^4 3^3 A^2, \tag{14}$$

which is of type (1). Hence there is a bound for the rank via $\mathfrak{G}/2\mathfrak{G}$. Selmer found evidence that the difference between the two bounds is always even.

The most plausible way in which even numbers naturally arise is as the rank of a skew-symmetric matrix. In the mean-time, for any elliptic curve $\mathcal{C}$ Tate and Shafarevich had constructed a group, usually denoted by Ш, which gives the obstruction to a local-global principle. With several years' hard work I managed to find([9]) a skew-symmetric form on

([8]) Selmer exemplifies the early connection between number theory and computers. He has been called the father of computer science in Norway.

([9]) Cassels (1962). For generalizations see Tate (1962) and Milne (1986).

Ш whose kernel is the set of infinitely divisible elements. If one assumes that the only infinitely-divisible element is 0, this implies Selmer's conjectures.

We now come to the celebrated conjectures of Birch and Swinnerton-Dyer (1965), which arose from a search for some quantitative local-global principle for elliptic curves analogous to Siegel's quantitative local-global principle for quadratic forms. We were then colleagues in Cambridge, so I had a ring-side seat. As I remember it, the conjectures were gradually refined from a vague hunch by the interaction of computer experiments and theoretical considerations. The final conjecture connects a quantity depending on $\mathfrak{G}$ with the behaviour of an L-function attached to the curve. There was one missing piece: a number occurred which could not be accounted for. For most of the curves investigated this number was 1: it was always an integer, though there was no obvious reason why this should happen, and indeed it was always a square. The skew-symmetric form on Ш ensures that its order, if finite, is a square, so a natural guess was that this was the unidentified number. This was then supported by other evidence.

The Birch-Swinnerton-Dyer conjectures have set the agenda for much that has happened since. They have been widely generalized and tested, but only recently are proofs beginning to emerge.

One check on the truth of these conjectures is that they can predict the existence of points on curves which are not immediately apparent. Andrew Bremner observed that for prime p the Birch-Swinnerton-Dyer conjectures imply the existence of a generator of infinite order on the curves

$$Y^2 = X(X^2 + p) \qquad p \equiv 5 \pmod 8 \tag{15}$$

in addition to the point $(0, 0)$ of order 2, and he devised a technique to look for it. I was roped in to the search. We looked at $p \leqslant 1000$. For $p = 877$ the generator is $(u/v, r/s)$, where

$$\begin{cases} u = 37549452812716219310550406994209279234620 1, \\ v = \quad 62159877768715054254632207806972380441 00, \\ r = 25625626798892680938877683404551308964866915320435660346478694 9, \\ s = \quad 49007802321978758895980293399592892509606161647077997926100 0. \end{cases} \tag{16}$$

(Cassels and Bremner (1984)). In fact the existence of these generators is predicted by the Selmer conjecture, but Birch-Swinnerton-Dyer even predicts the size of the numbers (the *height* of the generator). Bremner (1988) and Bremner and Buell (1993) have carried the search further. Other examples of generators of great height have been given by Zagier.

We now change the subject to cubic surfaces $f(X, Y, Z, T) = 0$, where the cubic form f has rational coefficients. Mordell conjectured that there is a local-global principle and there were proofs for surfaces of some special types. For example, Selmer showed that the principle holds for «diagonal» cubic surfaces

$$aX^3 + bY^3 + cZ^3 + dT^3 = 0, \tag{17}$$

provided that ab/cd is a rational cube. However Swinnerton-Dyer gave an ingenious example of a form for which the local-global principle fails. There remained the possibility that the principle holds for all «diagonal» surfaces, whether or not they satisfy Selmer's condition. This seemed to me unlikely. At that time Mike Guy was my research student and already hooked on computers. I suggested that he look for a likely counter-example. So he programmed the computer to look for all surfaces (17) with coefficients up to 50 which (i) are everywhere locally soluble and (ii) have no solutions with the variables up to 50. In the primitive state of computers at the time this was no small achievement. There was a snag. The demand for computer time was high and programs had to be fed in directly, no on-line facilities. Consequently keen users, particularly if they were junior, did all their work at night. I am a normal sort of chap who works by day. Although I knew at second hand that Guy had a list of potential counter-examples, I could not get them. This deadlock persisted for some time. Finally, John Horton Conway, who works both by day and by night, gave me, not the whole list, but what he thought was the simplest specimen

$$5X^3 + 12Y^3 + 9Z^3 + 10T^3 = 0. \tag{18}$$

I took this with me on holiday. Fortunately from this point of view, the weather was atrocious: I had plenty of time to think, and was able to prove that there is no global point (Cassels and Guy (1966)).

Note that although the computer was crucial to the discovery, it does not affect the logical status of the result. This would have been exactly the same if the Angel Gabriel had appeared to me in a dream and said «why not try (18)?». In fact the simplest counter-example on Guy's list is not (18) but

$$X^3 + 4Y^3 + 10Z^3 + 25T^3 = 0, \tag{19}$$

which involves only two primes, while the coefficients of all the other listed forms contain at least three. Bremner (1978) showed by a rather different argument that there are no global points on (19).

Manin (1974) has produced a rather sophisticated invariant, which he conjectures is the only obstruction to the local-global principle for cubic surfaces. For «diagonal» surfaces Colliot-Thélène, Kanevsky and San-

suc (1987) give supporting computational evidence. They confined attention to diagonal surfaces because it is difficult to teach even a powerful modern machine to compute Manin's invariant for general forms.

We conclude with a problem of a very different type. Let $f(X) \in \mathbb{C}[X]$, $g(Y) \in \mathbb{C}[Y]$ be polynomials with complex coefficients. When does

$$f(X) - g(Y) \tag{20}$$

have a factor in $\mathbb{C}[X, Y]$? A familiar example is $f = g$, when there is the factor $X - Y$, or, more generally, $f(X) = H(F(X))$, $g(Y) = H(G(Y))$ with factor $F(X) - G(Y)$. Davenport, Lewis and Schinzel (1961) found another example. Let

$$T_4(Z) = \cos(4 \arccos Z) = 8Z^4 - 8Z^2 + 1 \tag{21}$$

be the 4th Chebyshev polynomial. Then

$$T_4(X) + T_4(Y) = \tag{22}$$

$$\{\sqrt{2}(2X^2 + 2Y^2 - 1) + 4XY\}\{\sqrt{2}(2X^2 + 2Y^2 - 1) - 4XY\}.$$

More generally, take $f = T_4(F(X))$, $g(Y) = -T_4(G(Y))$. Davenport, Lewis and Schinzel were rash enough to suggest that this exhausts the possibilities.

It seemed a good idea to look at the associated Riemann surfaces. The equation $f(X) = Z$ gives a covering of the Riemann sphere of Z, and similarly for $g(Y) = Z$. We can combine the coverings by looking at the pair (X, Y) over Z, where

$$f(X) = g(Y) = Z. \tag{22}$$

Then $f(X) - g(Y)$ factors precisely when the covering (22) consists of more than a single piece. This reduces the problem to a purely combinatorial one about the behaviour at the ramification points. The most promising case is when f, g have the same degree n (say). Mike Guy programmed this on EDSAC2, the successor to EDSAC, and he got as far as $n = 12$ before EDSAC2 was killed: its memory was required for the newly-acquired TITAN ([10]), which was much more powerful but which, as happened in those days, required programming in an entirely different way.

Guy found new solutions of the combinatorial problem for $n = 7, 11$.

([10]) TITAN was a cut-down and modified ATLAS, a computer much ahead of its time. It is a tragedy of British post-war economic policy that resources were refused to develop it, but lavished on such clearly uneconomic prestige projects as Concorde.

This left the question of finding the corresponding polynomials f and g. We were still pondering how to do this when Bryan Birch visited Cambridge. Guy told him the problem over lunch and Bryan then and there found the polynomials for $n = 7$. Over dinner he found those for $n = 11$. The polynomials for $n = 7$ contain a parameter t. With

$$\lambda = (1 + \sqrt{-7})/2, \qquad \mu = (1 - \sqrt{-7})/2 \tag{23}$$

put

$$\begin{cases} f(X) = X^7 - 7\lambda t X^5 + (4 - \lambda) t X^4 + (14\lambda - 35) t^2 X^3 - \\ \qquad - (8\lambda + 10) t^2 X^2 + ((3 - \lambda) t^2 + 7(3\lambda + 2) t^3) X, \\ g(Y) = Y^7 - 7\mu t Y^5 + (4 - \mu) t Y^4 + (14\mu - 35) t^2 Y^3 - \\ \qquad - (8\mu + 10) t^2 Y^2 + ((3 - \mu) t^2 + 7(3\mu + 2) t^3) Y + 7t^3. \end{cases} \tag{24}$$

Then

$$f(X) - g(Y) = U(X, Y) V(X, Y), \tag{24}$$

where

$$\begin{cases} U = X^3 + \lambda X^2 Y - \mu X Y^2 - Y^3 - (3\lambda + 2) t X + (3\mu + 2) t Y + t, \\ V = X^4 - \lambda X^3 Y - X^2 Y^2 - \mu X Y^3 + Y^4 + 2(\mu - \lambda) t X^2 - 7t X Y + \\ \qquad + 2(\lambda - \mu) t Y^2 + (3 - \lambda) t X - (3 - \mu) t Y - 7t^2. \end{cases} \tag{25}$$

The formulae for $n = 11$ are similar, but more complicated and without a parameter.

I talked about this at the 15th Scandinavian Congress (Cassels (1968)), where I learned that Tverberg had found the 7 factorization from a different point of view. Subseqently it has been shown that the classification of the simple finite groups (!) implies that there are essentially only finitely many further such curious factorizations, and they have all been determined, at least as sphere coverings (Feit (1980)).

I learned only recently that Birch had already been considering coverings of the Riemann sphere in a different context. They are also the starting point of an unpublished research program of Grothendieck which has attracted much attention recently and which was the theme of a recent conference at Luminy. [See: *The Grothendieck theory of dessins d'enfants* (ed. L. Schneps), LMS Lecture Notes, **200** (1994).]

I hope that I have now said enough to convince you of my thesis that the theory of numbers is an experimental subject and that nowadays the sensible way to experiment is usually on a computer.