BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 41, Number 1, Pages 43–57
S 0273-0979(03)00993-5
Article electronically published on September 5, 2003

CATALAN'S CONJECTURE: ANOTHER OLD DIOPHANTINE PROBLEM SOLVED

TAUNO METSÄNKYLÄ

ABSTRACT. Catalan's Conjecture predicts that 8 and 9 are the only consecutive perfect powers among positive integers. The conjecture, which dates back to 1844, was recently proven by the Swiss mathematician Preda Mihăilescu. A deep theorem about cyclotomic fields plays a crucial role in his proof.

Like Fermat's problem, this problem has a rich history with some surprising turns. The present article surveys the main lines of this history and outlines Mihăilescu's brilliant proof.

1. INTRODUCTION

Catalan's Conjecture in number theory is one of those mathematical problems that are very easy to formulate but extremely hard to solve. The conjecture predicts that 8 and 9 are the only consecutive perfect powers, in other words, that there are no solutions of the Diophantine equation

$$x^u - y^v = 1 \qquad (x > 0,\ y > 0, \quad u > 1,\ v > 1) \tag{1.1}$$

other than $x^u = 3^2,\ y^v = 2^3$.

This conjecture was received by the editor of the *Journal für die Reine und Angewandte Mathematik* from the Belgian mathematician Eugène Catalan (1814–1894). The journal published it in 1844 [CAT]. Catalan, at that time a teacher at l'École Polytechnique de Paris, had won his reputation with a solution of a combinatorial problem. The term *Catalan number*, still in use, refers to that problem. As to the equation (1.1), Catalan wrote that he "could not prove it completely so far." He never published any serious partial result about it either.

The conjecture became a challenge for mathematicians, and some interesting results about particular cases of the equation appeared soon. However, during the first hundred years or so, all results were of a more or less isolated nature. After that, towards the end of the 1950s, there were several remarkable ideas evolving almost simultaneously, and in the 1970s the study was electrified by a result reducing the problem to a finite computation. Yet it appeared that the required computational work was all too big to be feasible. From that time on, the main direction of the study was in the efforts to diminish the amount of that work.

This was the situation in 2002 when Preda Mihăilescu, a mathematician practically unknown to the experts in this area, turned up with a complete proof of the

Received by the editors March 5, 2003, and, in revised form, July 14, 2003.
2000 *Mathematics Subject Classification.* Primary 11D41, 00-02; Secondary 11R18.
Key words and phrases. Catalan's Conjecture, Diophantine equations of higher degree, cyclotomic fields, research exposition.

conjecture. Surprisingly, his proof has very little to do with computation, making instead use of deep theoretical results, notably from the theory of cyclotomic fields.

Mihăilescu, born in 1955 in Romania, received his mathematical education at the ETH Zürich. He has worked in the machine and finance industry but is now doing research in Germany at the University of Paderborn.

The present article describes briefly the landmarks in the history of the work on Catalan's problem and outlines Mihăilescu's brilliant solution.

2. Early and not-so-early history

About 100 years before Catalan sent his letter to Crelle, Euler had proven that 8 and 9 are the only consecutive integers among squares and cubes, that is, the only solution of the Diophantine equations

$$x^3 - y^2 = \pm 1 \qquad (x > 0,\ y > 0). \tag{2.1}$$

Euler's proof is ingenious but somewhat tedious. Among other things, it resorts to the method of infinite descent due to Fermat.

As a background to (1.1) it is illuminating to look at how the special equations (2.1) are solved by current methods of algebraic number theory. Let (x, y) be a solution. Take first the equation with the plus sign and write it in the ring $\mathbb{Z}[i]$ of Gaussian integers as

$$x^3 = (y + i)(y - i). \tag{2.2}$$

Since $\mathbb{Z}[i]$ is a unique factorization ring, we may speak about the GCD of its elements. Let d be the GCD of $y + i$ and $y - i$ (unique up to a unit factor). The equations $y + i = d\lambda,\ y - i = d\mu$ imply that $d|2$. On the other hand, d divides x and x must be odd, since $y^2 \equiv 0$ or $1 \pmod 4$. Hence d is a unit, that is, one of the numbers $\pm 1, \pm i$.

We have $y + i = d(a + bi)^3$ with $a, b \in \mathbb{Z}$. But d is anyway a 3rd power in $\mathbb{Z}[i]$ and so may be ignored here. Looking separately at the real and imaginary parts in the equation $y + i = (a + bi)^3$, we find that $y = 0$ (and $x = 1$), a contradiction. Thus there is no solution.

The second equation will be written in a similar form,

$$x^3 = (y + 1)(y - 1).$$

The GCD of $y + 1$ and $y - 1$ is 1 or 2. In the former case we see that 2 would be a difference of two cubes, which is impossible. The latter case correspondingly leads to the equations

$$a^3 - 2b^3 = \pm 1.$$

Hence the number $a - b\alpha$, with $\alpha = \sqrt[3]{2}$, is a unit in $\mathbb{Z}[\alpha]$, the ring of integers in the real cubic field $\mathbb{Q}(\alpha)$. The units of this ring are, up to sign, powers of the single unit $1 + \alpha + \alpha^2$. With some work one finds that $|a - b\alpha|$ can only be the zeroth power, so that $a = \pm 1$ and $b = 0$. Returning to the original equation we get the solution $x = 2$, $y = 3$.

To prove Catalan's Conjecture, it obviously suffices to consider the equation

$$x^p - y^q = 1 \qquad (x > 0,\ y > 0), \tag{2.3}$$

where p and q are different primes.

The case of $q = 2$ was solved in 1850 by V.A. Lebesgue [LEB] (not to be confused with his more famous namesake). At that time one already knew the arithmetic of Gaussian integers, so that the equation could be treated in the form

$$x^p = (y + i)(y - i),$$

analogous to (2.2). The GCD of $y + i$ and $y - i$ is again a unit, but this time it cannot be easily ignored. Hence we have a pair of equations,

$$y + i = i^s(a + bi)^p, \qquad y - i = (-i)^s(a - bi)^p,$$

where $s \in \{0, 1, 2, 3\}$. From these, y can be eliminated in several ways, and the equations so obtained lead to a contradiction. Thus the equation $x^p - y^2 = 1$ has no solution.

How about the case of $p = 2$ in (2.3)? This remained a mystery for a long time. As late as 1961 a result was published proving that a possible solution of $x^2 - y^q = 1$ necessarily has $x > 10^{3 \cdot 10^9}$. The news that a Chinese mathematician, Chao Ko, had just proven this equation insolvable had not yet reached the mathematical community. The proof became known in 1964 when it appeared in *Scientia Sinica* [KO].

In 1976 E.Z. Chein [CHE] published a new, very elegant proof. It is based on a result by T. Nagell [NAG] stating that the solution (x, y) must satisfy $2|y$ and $q|x$. Considering the equation in the form

$$(x + 1)(x - 1) = y^q,$$

Chein concluded that the GCD of $x + 1$ and $x - 1$ is 2, so that there are coprime integers a and b, with a odd, satisfying the equations

$$x + 1 = 2a^q, \qquad x - 1 = 2^{q-1}b^q \tag{2.4}$$

or, alternatively, similar equations with $x+1$ and $x-1$ interchanged. If $q > 3$, then (2.4) can be shown to imply a condition

$$(ha)^2 + b^2 = (a^2 - b)^2,$$

where $h^2 = a^2 - 2b$, and the alternative equations yield a similar condition. These are both Pythagorean type equations, and thus their complete solution is known. From this it follows that x and y fail to exist (for $q > 3$).

More details about the above solutions can be found in Paulo Ribenboim's nice monograph [RIB]. That book gives a comprehensive history of Catalan's Conjecture until 1994.

3. Cassels and Case I

From now on it is convenient to consider Catalan's equation in the form

$$x^p - y^q = 1 \qquad (xy \neq 0,\ p, q \text{ different odd primes}). \tag{3.1}$$

Thus, unless otherwise stated, negative integers x, y are allowed as well.

By way of "multiplicatization" of the equation, rewrite it as

$$(x - 1)\frac{x^p - 1}{x - 1} = y^q.$$

What is the GCD of the two factors on the left hand side? By considering the identity $x^p = ((x - 1) + 1)^p$ one easily finds that there are two possibilities: the GCD is either 1 or p.

A similar situation occurs in the study of the Fermat equation $x^p + y^p = z^p$, where the left hand side can be split into the product of $x+y$ and $(x^p+y^p)/(x+y)$. Here, too, the GCD of these factors is 1 or p, and this leads to *Case I* and *Case II* of the problem, respectively. Historically, Case I was "easier", and many people felt optimistic that the classical approach in this case might eventually prove successful. In the solution by Andrew Wiles this kind of classification played no role.

For the equation (3.1) we may similarly speak about Case I and Case II according to the value of the GCD above. In Case I, when the GCD equals 1, we obtain the equations

$$x - 1 = a^q, \qquad \frac{x^p - 1}{x - 1} = b^q, \qquad y = ab,$$

where a and b are coprime and not divisible by p. In 1960 J.W.S. Cassels [CAS] showed that these equations yield a contradiction. His method is elementary, a striking mixture of divisibility relations and inequalities. A different proof was later discovered by S. Hyyrö [HY2], who somewhat amazingly could apply his results about the Diophantine equation $ax^n - by^n = z$.

This means that we are left with Case II. In particular, one of the two numbers $x-1$ and $(x^p-1)/(x-1)$ contains p just in the first power. But this number cannot be $x-1$, since in that case x^p-1 would only be divisible by p^2. Therefore we have the equations

$$x - 1 = p^{q-1}a^q, \qquad \frac{x^p - 1}{x - 1} = pb^q, \qquad y = pab, \tag{3.2}$$

where again a and b are coprime and p does not divide b (but p may divide a). Analogous equations follow from the factorization of x^p into the product of $y+1$ and $(y^q+1)/(y+1)$. In particular, y is divisible by p and x is divisible by q.

From the last result it follows, by the way, that there cannot be *three* consecutive perfect powers – a good exercise for the interested reader.

Cassels' theorem was one of the first general results about Catalan's equation (3.1). It gave a significant impulse to the study of this equation.

4. Can the problem be solved by a computer?

Around the middle of the last century Catalan's Conjecture began to interest people working on Diophantine analysis. One of their early observations was that the number of solutions (x, y) to (3.1), for fixed exponents p and q, is at most finite. This is a consequence of a general theorem about integer points on a curve, published by C.L. Siegel in 1929. A result by H. Davenport and K.F. Roth from 1955 further allowed one to derive an explicit – though enormous – upper bound for that number, as shown in [HY1]. (For other results about the number of solutions, see the introductory section in [TIJ].)

A turning point in this direction came in the seventies, some time after Alan Baker had obtained his fundamental estimates for linear forms of logarithms. To give an idea of Baker's result, let

$$\Lambda = b_1 \log r_1 + \cdots + b_n \log r_n,$$

where the b_j are integers and the r_j are positive rational numbers. Define the height of a rational number $r = s/t$ (in lowest terms) as $\log \max(|s|, |t|)$ and let

$B = \max(|b_1|, \ldots, |b_n|)$. Suppose that $\Lambda \neq 0$. Baker proved an inequality of the form

$$|\Lambda| > \exp(-A \log B),$$

where A is an explicitly computable positive number depending on n and on the heights of $r_1, \ldots, r_n$.

This result, or in fact a slight refinement, was used by Robert Tijdeman [TIJ] to bound a solution (x, y, p, q) (with x, y positive) of Catalan's equation from above. The strategy is to find linear forms Λ depending on this solution in a special way: an upper bound for $|\Lambda|$ implied by (3.1) should be sufficiently close to Baker's lower bound. Tijdeman's clever choices were

$$\Lambda_1 = q \log q - p \log p + pq \log \frac{pa}{qa'} = \log \frac{(x-1)^p}{(y+1)^q},$$
$$\Lambda_2 = q \log q + p \log \frac{p^{q-1}a^q + 1}{q^q a'^q} = \log \frac{y^q + 1}{(y+1)^q},$$

where a is defined by the equation $x - 1 = p^{q-1}a^q$ in (3.2) and a' is defined by the analogous equation $y + 1 = q^{p-1}a'^p$. Since

$$(x-1)^p < x^p = y^q + 1 < (y+1)^q,$$

Λ_1 and Λ_2 are nonzero. A comparison of lower and upper bounds of $|\Lambda_1|$ leads to an inequality between p and q, and the same happens with $|\Lambda_2|$. These inequalities are indeed sharp enough to yield, on eliminating q, a condition

$$p < c_1 (\log p)^{c_2},$$

where c_1 and c_2 are constants. This requires that $q < p$, but in the case $q > p$ a similar argument gives a corresponding condition for q.

It follows that the exponents p and q are bounded from above, and the bound is independent of x and y. So the nature of our problem changed dramatically: there is only a finite number of solutions (x, y, p, q), and still more, one can compute explicit bounds for the unknowns.

Indeed, the constants c_1, c_2 above are effective. The first explicit computations produced astronomical upper bounds for p and q, but refinements of the method have given estimates of more modest size. By the best result, the upper bound for $\max(p, q)$ is about $8 \cdot 10^{16}$ (for this and similar results, see Mignotte's article [MI2]). This limit is of course still far beyond anything useful for practical purposes.

Note that when restricting to positive x and y above we did not lose anything, since (3.1) may also be written in the form

$$(-y)^q - (-x)^p = 1. \tag{4.1}$$

A very instructive exposition of a Tijdeman-type reasoning based on more recent estimates of logarithmic forms (in the special cases of interest here) can be found in Yuri Bilu's article [BIL].

What happens, by the way, when our equation is replaced by $x^p - y^q = c$, where $c > 1$ is a given integer? For fixed p and q, Siegel's theorem again implies that the number of solutions (x, y) is finite. But allowing the exponents to vary makes the situation much more complicated: we do not know whether the number of solutions (x, y, p, q) is finite, not to mention any effective upper bound.

5. The Wieferich pairs

The preceding results prompted a search for further restrictions on putative solutions.

Recall that we transformed our equation $x^p - y^q = 1$ in (3.2) in the form

$$\frac{x^p - 1}{x - 1} = pb^q \qquad (p \nmid b).$$

This suggests a traditional approach of factorizing the left hand side in $\mathbb{Z}[\zeta]$, the ring of integers in the pth cyclotomic field $\mathbb{Q}(\zeta)$ ($\zeta = e^{2\pi i/p}$). Combining this procedure with the observation that

$$p = \Big[\frac{x^p - 1}{x - 1}\Big]_{x=1} = \prod_{k=1}^{p-1}(1 - \zeta^k),$$

we obtain the equation

$$\prod_{k=1}^{p-1} \frac{x - \zeta^k}{1 - \zeta^k} = b^q.$$

Write $x-\zeta^k = (x-1)+(1-\zeta^k)$ and notice that $x-1$ was found to be divisible by p (see (3.2)). It follows that the quotients $(x-\zeta^k)/(1-\zeta^k)$ are in $\mathbb{Z}[\zeta]$. Unfortunately, this ring is no UFD in general. To restore good factorization properties we thus have to replace our numbers with the ideals they generate. Now it is not difficult to show that the principal ideals $\langle (x - \zeta^k)/(1 - \zeta^k)\rangle$ are pairwise coprime. Hence each of them is a qth power of some ideal, in particular

$$\langle \frac{x - \zeta}{1 - \zeta} \rangle = J^q, \tag{5.1}$$

where J is a nonzero ideal of $\mathbb{Z}[\zeta]$.

All this is analogous to Kummer's classical work on the Fermat equation and was published by K. Inkeri [IN2] in 1990. In his article Inkeri went on in the same spirit by making the assumption that the class number of the field $\mathbb{Q}(\zeta)$ is prime to q. Then, along with J^q the ideal J has to be principal, say $J = \langle\gamma\rangle$, and (5.1) implies an equation

$$\frac{x - \zeta}{1 - \zeta} = \epsilon\gamma^q, \tag{5.2}$$

where ϵ is a unit in $\mathbb{Z}[\zeta]$.

There are infinitely many units in this ring, but one can overcome this obstacle. Indeed, consider (5.2) together with its complex conjugate and use the fact that ϵ and $\overline{\epsilon}$ differ by a factor which is a root of unity. In this manner Inkeri was able to obtain, after some manipulation, the result that $q^2|x$. Remember that we had $q|x$ by Cassels. The strengthening here is based on the *Rule of Lifting the Exponent*, which every student of number theory meets in this elementary form: if $a^q \equiv b^q \pmod q$, then $a^q \equiv b^q \pmod{q^2}$. We will come across this rule several times in the sequel.

On rewriting the first equation in (3.2) as

$$x = (p^{q-1} - 1)a^q + a^q + 1,$$

Inkeri drew the further consequence that q^2 divides $p^{q-1}-1$. By (4.1) the roles of p and q can be interchanged, and thus we have a strange pair of congruences,

$$p^{q-1} \equiv 1 \pmod{q^2}, \qquad q^{p-1} \equiv 1 \pmod{p^2}, \tag{5.3}$$

provided the class numbers of the pth and qth cyclotomic fields behave well; that is, the former class number is prime to q and the latter prime to p. A pair of odd primes p, q satisfying these congruences is now called a (*double*) *Wieferich pair*. This name has its origin in the history of the Fermat problem: A. Wieferich showed in 1909 that the solvability of the equation $x^p + y^p = z^p$ in Case I requires that $2^{p-1} \equiv 1 \pmod{p^2}$. Such primes p, *Wieferich primes*, turned out to be extremely rare. In fact, there are just two Wieferich primes known, 1093 and 3511, and the next one, if it exists, must exceed $1.25 \cdot 10^{15}$ (see [DIL], [KNR]).

Also the Wieferich pairs are very exceptional. The first pair that was found is (83, 4871), and only five further pairs are known ([MI2], [KER]).

The conditions (5.3) together with existing class number tables were used to rule out a large family of p and q in the possible solutions of Catalan's equation. This method grew still more efficient as people found ways to modify and relax those class number conditions ([MI1], [SCH], [STE]; see also [IN1]).

The most dramatic progress in this direction happened in 1999, when Preda Mihăilescu [M1] proved, as a *prelude* to his coming *opus magnum*, that the congruences (5.3) in fact hold without any class number condition.

6. Annihilators as key actors

The critical point requiring a class number condition above was passing from the ideal equation (5.1) back to an equation between numbers. Mihăilescu's idea was to do this transition in a different way, by means of annihilators of ideals.

Annihilating an element of a group means mapping it to the neutral element e; annihilating the whole group means mapping all of it to e. In the case of the (ideal) class group of a number field, e is the class of principal ideals. Thus the annihilator of a (nonzero) ideal is a slightly loose expression for a map sending this ideal to a principal ideal. The reader should excuse this long explanation, but annihilators do really play a vital role in the proof of Catalan's Conjecture.

If θ annihilates the ideals of the ring $\mathbb{Z}[\zeta]$, the equation (5.1) implies that

$$\left(\frac{x-\zeta}{1-\zeta}\right)^{\theta} = \epsilon\gamma^q, \tag{6.1}$$

where $\epsilon \in \mathbb{Z}[\zeta]^\times$ as before and $\gamma \in \mathbb{Q}(\zeta)$ is defined by $J^\theta = \langle\gamma\rangle$.

In his first article Mihăilescu [M1] chose a classical annihilator given by the so-called Stickelberger relation. A calculation similar to Inkeri's but technically more involved then yields the congruences

$$x \equiv 0, \qquad p^{q-1} \equiv 1 \qquad (\mathrm{mod}\ q^2). \tag{6.2}$$

By the symmetry mentioned above, we also have

$$y \equiv 0, \qquad q^{p-1} \equiv 1 \qquad (\mathrm{mod}\ p^2). \tag{6.3}$$

We repeat that (6.2) and (6.3) must be satisfied by any solution (x, y, p, q) of Catalan's equation (3.1). In particular, the exponents p, q form a Wieferich pair.

These are extremely efficient conditions in eliminating possible solutions of (3.1). By combining them with suitable inequalities for p and q, obtained by Tijdeman-type methods, Mignotte and others showed by computation that $\min(p, q) > 10^7$; see [MI2]. There are subsequent improvements of this bound, but after all, the bound is still so far from the upper bound mentioned in Section 4 that the problem remains "cryptographically secure", as Mihăilescu once put it.

The main value of (6.2) and (6.3) is of a theoretical kind, however, as we shall soon see.

In his eventual proof of the conjecture, Mihăilescu [M2] looks at the equation (6.1) from a totally new point of view. Instead of trying to push the unit ϵ aside, he just focuses on this unit or, in fact, on the whole group of units. He explores the information provided for this group by (6.1) through different annihilators θ and finds one unexpected property of that group. On showing that such a property is absurd, he comes to a contradiction proving the conjecture.

Suitable means for this program are offered by a deep result about annihilators called Thaine's theorem. This result in its general form concerns real abelian fields; in the next section it will be quoted in the case of interest to us.

For readers familiar with abelian fields we would like to point out that, along with Thaine's theorem, Mihăilescu's proof may be seen as belonging to the "plus-part" of the cyclotomic theory. This is in contrast to the initial step [M1] receiving a crucial ingredient from the "minus-part" of the theory.

7. Special annihilators

The natural setting for the following considerations is the real cyclotomic field $K = \mathbb{Q}(\zeta) \cap \mathbb{R}$. This is an extension of degree $m = (p-1)/2$ over the rationals, generated by $\rho = \zeta + \zeta^{-1}$, for example. Its ring of integers is $\mathbb{Z}[\rho]$. The group of units of this ring, $E = \mathbb{Z}[\rho]^\times$, is an infinite abelian group generated by -1 and by $m-1$ torsion-free units, the *fundamental units* of K. These are in general truly hard to find, but as a kind of replacement, consider the units

$$\frac{\sin(l\pi/p)}{\sin(\pi/p)} = \frac{\zeta^{l/2} - \zeta^{-l/2}}{\zeta^{1/2} - \zeta^{-1/2}} \qquad (l = 2, \dots, m)$$

(note that $\zeta^{1/2} = -\zeta^{(p+1)/2}$). Together with -1 these units generate an important subgroup of E of finite index. Denote it by C. The elements of C are called *cyclotomic* or *circular units*.

There is a surprising link, discovered by Kummer, between those unit groups and the class group $H(K)$ of K. In fact, the index $[E : C]$ equals the class number $h_K = |H(K)|$. This result has been extended and sharpened in several ways. A last step in this development is Thaine's theorem [THA] relating annihilators of units to those of ideals. Before going into details we have to introduce the annihilators more precisely.

The field K is a Galois extension of $\mathbb{Q}$, its group G consisting of the automorphisms $\tau_1, \dots, \tau_m$ defined by $(\zeta + \zeta^{-1})^{\tau_c} = \zeta^c + \zeta^{-c}$. It is natural to introduce a larger set of maps, the group ring

$$\mathbb{Z}[G] = \{ \textstyle\sum_{c=1}^{m} n_c \tau_c \mid n_c \in \mathbb{Z} \ (c = 1, \dots, m) \}.$$

The Galois action of G in the field K induces a $\mathbb{Z}[G]$-module structure on $K^\times$, the multiplicative group of K, by the rule

$$\gamma^{n_1\tau_1+\cdots+n_m\tau_m} = (\gamma^{n_1})^{\tau_1}\cdots(\gamma^{n_m})^{\tau_m} \qquad \forall\ \gamma \in K^\times.$$

In particular, the groups E and C become submodules of $K^\times$. Observe that it was convenient to adopt the exponential notation for the module action since the groups are multiplicative.

The ring $\mathbb{Z}[G]$ also acts on the ideal group of K and, along with this, on the class group $H(K)$. Thus $H(K)$ is a $\mathbb{Z}[G]$-module as well.

The domain where our annihilators will be chosen from is $\mathbb{Z}[G]$.

For an abelian group A denote by $[A]_q$ the q-primary subgroup of A, that is, the subgroup consisting of elements with order a q-power. If A is a $\mathbb{Z}[G]$-module, then $[A]_q$ is a submodule.

Thaine's theorem for the field K and the (odd) prime q reads as follows: *If the degree $m = [K : \mathbb{Q}]$ is prime to q, then every annihilator $\theta \in \mathbb{Z}[G]$ of the group $[E/C]_q$ also annihilates the group $[H(K)]_q$.*

To be precise, the group of cyclotomic units used by Thaine is not quite the same as our C but instead a subgroup of C of index 2^{m-1}(cf. [LET]). This difference does not matter here (since q is odd) and will be ignored in the sequel.

Remark. Thaine's article was published in 1988, but the theorem was known to be true before this. In fact, R. Greenberg [GRE] had shown that the "main conjecture" of Iwasawa theory implies a conjecture by G. Gras saying that the groups $[E/C]_q$ and $[H(K)]_q$, regarded as $\mathbb{Z}_q[G]$-modules, have isomorphic Jordan–Hölder series. (Here $\mathbb{Z}_q$ stands for the q-adic integers.) The main conjecture was proven by B. Mazur and A. Wiles [MW] in 1984. See [MW], p. 214, and also K. Rubin's review of [THA].

However, Thaine's method of proof is more direct. A good exposition of this proof can also be found in L.C. Washington's monograph [WAS]. Anyway, the proof shows that behind this theorem – even in the present special case – there is a substantial amount of theory, including class field theory.

The first task now is to ensure that the condition $q \nmid m$ holds true in our situation. We argue indirectly. If $q|m$, then $p \equiv 1 \pmod q$ and hence, by the Rule of Lifting the Exponent, $p^q \equiv 1 \pmod{q^2}$. On the other hand, $p^q \equiv p \pmod{q^2}$ by (6.2). Consequently, $p \equiv 1 \pmod{q^2}$. Since q^2+1, $2q^2+1$ and $3q^2+1$ are not primes, it follows that $p > 4q^2$.

However, an argument based on linear forms of logarithms (see Section 4) shows that $p < 4q^2$ whenever $q > 28000$. Thus we have a contradiction unless the exponents p, q satisfy the inequalities $q < 28000$, $p > 4q^2$. But such exponents (and even a lot more) have already been excluded, as explained in Section 6. Bilu [BIL] reports on a tailor-made modification of the above argument ruling out just the last mentioned p, q. This required only 1 minute running time on a computer.

In the entire proof, this is the only place where a computer is needed. Note also that this is a place where one really uses a method originating in Tijdeman's distinguished work, although the actual result obtained by Tijdeman will not be needed.

There is more recent news that Mihăilescu has found a totally different approach for verifying the condition $q \nmid m$. This would apply the "minus-part" theory referred

to above and be free of any computer calculations. One important link to that work is an article by Y. Bugeaud and G. Hanrot [BH].

8. Proof of Catalan's Conjecture outlined

In this section we sketch Mihăilescu's proof up to a statement about qth powers in the field K. This statement, which might be called Mihăilescu's key theorem, will be proven in Section 9. Sections 10 and 11 will shed more light on some details omitted below.

Let (x, y) be a solution of Catalan's equation (3.1). As stated in (5.1), the principal ideal in $\mathbb{Z}[\zeta]$ generated by $(x-\zeta)/(1-\zeta)$ is a qth power of a nonzero ideal. The same is true for the complex conjugate ideal, and multiplying those two ideals we get

$$\Big\langle \frac{(x-\zeta)(x-\zeta^{-1})}{(1-\zeta)(1-\zeta^{-1})} \Big\rangle = (J\overline{J})^q,$$

an equation between real ideals. In particular, the ideal class of $J\overline{J}$ has order q or 1 in the group $H(K)$ and so belongs to the q-primary group $[H(K)]_q$.

Let $\theta \in \mathbb{Z}[G]$ annihilate the factor group E/C, so that $E^\theta \subseteq C$. Then Thaine's theorem implies (as will be shown in Section 10) that θ annihilates $[H(K)]_q$. It follows, as in Section 6, that

$$\left(\frac{(x-\zeta)(x-\zeta^{-1})}{(1-\zeta)(1-\zeta^{-1})} \right)^\theta = \epsilon\gamma^q, \tag{8.1}$$

where $\epsilon \in E$ and $\gamma \in K^\times$. Since γ is unknown anyway, it is sufficient to consider ϵ, and units related to it, up to a factor which is a qth power in $K^\times$. In what follows this is usually done without mentioning it specifically.

Since $\epsilon^\theta \in C$, the unit ϵ in (8.1) can itself be assumed to be in C. This step requires a delicate property of annihilators to be discussed in Section 10.

A trivial but important choice for θ above is the norm map $N = \sum_c \tau_c$ or an integral multiple of it. Indeed, the norm of any unit is ± 1. For a suitable $r \in \mathbb{Z}$, the element $\big((1-\zeta)(1-\zeta^{-1})\big)^{\theta - rN}$ is a cyclotomic unit, and (8.1) then implies that

$$\big((x-\zeta)(x-\zeta^{-1})\big)^{\theta-rN} \in \eta(K^\times)^q, \qquad \eta \in C. \tag{8.2}$$

Since $x \equiv 0 \pmod{q^2}$, by (6.2), we find that $\eta \equiv 1 \pmod{q^2}$ (remember: η up to a qth power). The cyclotomic units satisfying this condition are called *q-primary* for historical reasons. They constitute a subgroup of C denoted by C_q.

Let $\theta' \in \mathbb{Z}[G]$ be an annihilator of C_q. Then it follows from (8.2) that

$$\big((x-\zeta)(x-\zeta^{-1})\big)^{\theta\theta'-rN} \in (K^\times)^q. \tag{8.3}$$

From this relation Mihăilescu is able to draw the conclusion that $\theta\theta' - rN$ is divisible by q, that is,

$$\theta\theta' - rN = q\omega, \qquad \omega \in \mathbb{Z}[G]. \tag{8.4}$$

The result (8.4) is all we need. Turning now to the group E we find that *every unit* $\epsilon \in E$ satisfies the condition

$$\epsilon^{\theta\theta'} = \epsilon^{rN+q\omega} = \epsilon^{rN} = 1.$$

Recalling that $\epsilon^\theta \in C$, this suggests that θ' in fact annihilates more of C than just the subgroup C_q, in this way forcing C_q to be equal to C. This argument can

really be made rigorous. Thus we have the result that all cyclotomic units should be q-primary.

It is fairly easy to show that this is impossible. After that we are done.

9. Mihăilescu's theorem

Mihăilescu's key result that (8.3) implies (8.4) has the following precise formulation. We recall that x denotes an integer that makes up, together with some y, a supposed solution of Catalan's equation $x^p - y^q = 1$.

Theorem. *Assume that* $\theta = \sum_{c=1}^{m} n_c\tau_c \in \mathbb{Z}[G]$ *and*

$$\big((x-\zeta)(x-\zeta^{-1})\big)^{\theta} \in (K^{\times})^q. \tag{9.1}$$

If $\sum_{c=1}^{m} n_c \equiv 0 \pmod q$, *then each* n_c *is divisible by* q, *so that* $\theta = q\omega$ *with* $\omega \in \mathbb{Z}[G]$.

The claim in this theorem may look quite plausible, but as a matter of fact the result is astonishing. Something like this has probably been in the mind of the people studying the Fermat problem, but in that context nothing similar has been found (see also [M2], Appendix C).

Crucial for the proof is the fact that $|x|$ is big. A good estimate is $|x| > q^p$, proven by Hyyrö [HY1] in 1964. Hyyrö's article appeared in the Turku University series, not easily available, but other estimates, occasionally weaker but anyway sufficient, are derived in [M2] and [BIL].

We will present the main lines of the proof of the theorem. There is no particularly deep mathematics involved. The general setting of the proof shows some similarity to the ideas of Bugeaud and Hanrot [BH].

To simplify notation, extend the automorphisms τ_c to the whole cyclotomic field $\mathbb{Q}(\zeta)$. The Galois group G_0 of $\mathbb{Q}(\zeta)$ consists of the automorphisms $\sigma_1, \dots, \sigma_{p-1}$ with $\zeta^{\sigma_k} = \zeta^k$. Hence τ_c has exactly the extensions σ_c and σ_{p-c}. Let

$$\psi = \sum_{c=1}^{m} n_c(\sigma_c + \sigma_{p-c}) = \sum_{k=1}^{p-1} b_k\sigma_k \in \mathbb{Z}[G_0],$$

where $b_c = n_c = b_{p-c}\,(c = 1, \dots, m)$. Then

$$\big((x-\zeta)(x-\zeta^{-1})\big)^{\theta} = (x-\zeta)^{\theta}(x-\zeta^{-1})^{\theta} = (x-\zeta)^{\psi}.$$

By adding to ψ a suitable element of the form $q\psi_1$, we may suppose that the coefficients b_k are in the range $0, \dots, q-1$. We have to show that each b_k in fact vanishes.

By the assumption of the theorem, $\sum_{k=1}^{p-1} b_k = tq$ with $t \in \{1, \dots, p-1\}$ (excluding the trivial case of $t = 0$). Since x is fixed by the σ_k, we have $(1-\zeta/x)^{\psi} = x^{-tq}(x-\zeta)^{\psi}$. Then, by (9.1),

$$\prod_{k=1}^{p-1}\left(1-\frac{\zeta}{x}\right)^{b_k\sigma_k} = \left(1-\frac{\zeta}{x}\right)^{\psi} = \gamma^q \qquad (\gamma \in K^{\times}).$$

A careful reflection shows that the real number γ can be expressed by means of a binomial series as follows:

$$\gamma = \prod_{k=1}^{p-1}\left(1-\frac{\zeta^k}{x}\right)^{b_k/q} = \prod_{k=1}^{p-1}\sum_{\mu=0}^{\infty}\binom{b_k/q}{\mu}\left(-\frac{\zeta^k}{x}\right)^{\mu} = \sum_{\mu=0}^{\infty}\alpha_\mu(\psi)\big(\frac{1}{x}\big)^{\mu}.$$

The coefficients $\alpha_\mu = \alpha_\mu(\psi)$ of the series are of the form $\alpha_\mu = a_\mu/(\mu! q^\mu)$, where $a_\mu \in \mathbb{Z}[\zeta]$. Let $q^{E(\mu)}$ denote the exact power of q dividing $\mu! q^\mu$.

Consider the remainder term $\Omega = \sum_{\mu=t+1}^{\infty} \alpha_\mu x^{-\mu}$, where t is the integer defined above. The number

$$\beta = q^{E(t)} x^t \Omega$$

is an integer of the field $\mathbb{Q}(\zeta)$, that is, belongs to $\mathbb{Z}[\zeta]$. One can estimate $|\beta|$ by means of a standard expression for the remainder term of a Taylor series. Applying Hyyrö's bound $|x| > q^p$, one arrives at the result $|\beta| < 1$. (This is not quite straightforward. One needs for t the bound $t \le m$ which can be achieved by the clever trick of replacing $\sum_k b_k \sigma_k$ by $\sum_k (q - b_k)\sigma_k$, if necessary. The series itself may be inconvenient to handle, but it can be replaced by a simpler "dominating" series – an ingenious idea due to Bilu.)

The argument can be extended to the conjugates β^{σ_k}; they will also be less than 1 in absolute value. But such a situation is impossible for a nonzero algebraic integer, and hence $\beta = 0$. Thus, our Taylor series for γ reduces to a finite sum!

On the other hand, evaluating the numerator of α_t gives

$$a_t \equiv \left(-\sum_{k=1}^{p-1} b_k \zeta^k \right)^t \pmod{q}.$$

In conjunction with the equation $\beta = 0$ this yields the congruence $\sum_k b_k \zeta^k \equiv 0 \pmod{q}$. This in turn is possible only if every b_k vanishes, the result that was to be proven.

10. Annihilators revisited

Working through the details of the proof described in Section 8 requires a deeper study of the annihilators. This brings an interesting algebraic aspect to the proof.

As stated in Section 8, it is sufficient to consider the units $\epsilon \in E$ modulo a qth power or, put exactly, to replace ϵ by its coset ϵE^q in the group E/E^q. When a map $\theta = \sum_c n_c \tau_c \in \mathbb{Z}[G]$ operates on the latter group, it is not the coefficients n_c that matter but just their residues modulo q. Thus we will regard the coefficients n_c either as integers or as their residue classes mod q, according to the situation at hand. In the latter case we have

$$\theta = \sum_{c=1}^{m} n_c \tau_c \in \mathbb{F}_q[G], \qquad \mathbb{F}_q = \mathbb{Z}/q\mathbb{Z},$$

and the group E/E^q becomes a module over the ring $R = \mathbb{F}_q[G]$.

To have the group $[E/C]_q$ from Thaine's theorem join the game, we correspondingly introduce the group E/CE^q, an R-module as well. Moreover, when "measuring" the difference between the groups C and C_q, it is natural to use the R-module CE^q/C_qE^q. In this way we come to study the three annihilators (i.e., sets consisting of the elements of R annihilating the module in question)

$$A_1 = \mathrm{Ann}(E/CE^q), \quad A_2 = \mathrm{Ann}(CE^q/C_qE^q), \quad A_3 = \mathrm{Ann}(C_qE^q/E^q).$$

As annihilators of R-modules these are ideals of R.

What does the ring R look like? It is essential that $\mathbb{F}_q$ be a field and the order of the group G be prime to q, the characteristic of this field. This gives R and its ideals a transparent structure. (By Maschke's theorem, R is a semisimple $\mathbb{F}_q$-algebra and

as such is decomposable into a direct sum of fields; see, e.g., [COH]. The ideals are principal, generated by sums of the idempotents defining that decomposition.)

Also the R-modules appearing above behave neatly: they are cyclic. It is enough to check this for the module E/E^q, since the other modules are obtained from it by the procedure of forming submodules and factor modules. The cyclicity of E/E^q appears to be quite a deep fact, and it would take too much space to review its proof here.

Every cyclic R-module M is (non-canonically) isomorphic to $R/\operatorname{Ann}(M)$, as is easy to verify. This isomorphism plus some information about the ideals of R enables one to conclude that the ideals A_1, A_2, A_3 are pairwise coprime and

$$A_1A_2A_3 = \operatorname{Ann}(E/E^q) = RN,$$

the principal ideal generated by the norm. Here the second equality follows from the cyclicity of E/E^q.

Every ideal I of R is idempotent, in other words, coincides with its square. Thus an element of I can always be written as a product of any number of elements of I. This is a convenient property (called "delicate" in Section 8) of annihilators to be used several times in the proof.

As a first illustration, let us show how one argues at the beginning of the proof that every $\theta \in A_1$ annihilates $[H(K)]_q$. Simply write $\theta = \theta_1 \cdots \theta_z$, where the θ_j belong to A_1 and z is defined by $|[E/C]_q| = q^z$. By the definition of A_1 we have $E^{\theta_j} \subseteq CE^q$ and so $E^\theta \subseteq CE^{q^z}$. Now let $\epsilon C \in [E/C]_q$. Then $\epsilon^\theta = \eta\epsilon_1^{q^z}$ with $\eta \in C$ and $\epsilon_1 \in E$, $\epsilon_1 C \in [E/C]_q$. It follows that $\epsilon^\theta C = (\epsilon_1 C)^{q^z} = C$. Consequently, θ annihilates the group $[E/C]_q$, and the assertion is indeed a consequence of Thaine's theorem.

Secondly, look at the equation (8.1) for $\theta \in A_1$. Write $\theta = \theta_1\theta_2$ with θ_1, θ_2 in A_1. Then the right hand side of (8.1) assumes the form

$$(\epsilon_1\gamma_1^q)^{\theta_2} = \epsilon_1^{\theta_2}(\gamma_1^{\theta_2})^q = \epsilon_2\gamma_2^q,$$

where $\epsilon_1 \in E$, $\epsilon_2 \in C$ and $\gamma_1, \gamma_2 \in K^\times$. Hence the unit ϵ in (8.1) can be chosen from C as claimed.

Once the reasoning outlined in Section 8 is carried through in a precise form, the relation corresponding to (8.3) says that

$$\big((x-\zeta)(x-\zeta^{-1})\big)^{\theta_1\theta_3 - rN} \in (K^\times)^q \tag{10.1}$$

for any $\theta_1 \in A_1$ and $\theta_3 \in A_3$, where $r \in \mathbb{F}_q$ is so chosen that the map $\theta_1\theta_3 - rN = \sum_c n_c\tau_c$ satisfies the condition $\sum_c n_c = 0$. (One should in fact regard $\theta_1\theta_3 - rN$ in (10.1) as lifted from $\mathbb{F}_q[G]$ to $\mathbb{Z}[G]$.) Mihăilescu's theorem then tells us that $\theta_1\theta_3 - rN = 0$. Consequently, $A_1A_3 \subseteq RN$. Noting that $RN = A_1A_2A_3$ and A_1, A_2, A_3 are pairwise coprime, we deduce that $A_2 = \langle 1\rangle$. From this it follows, by the definition of A_2, that $C = C_q$.

11. A contradiction, finally

The equality $C = C_q$ means that every cyclotomic unit in K, when regarded modulo q^2, is the qth power of some nonzero integer of K.

We will need the notion of cyclotomic units in the whole field $\mathbb{Q}(\zeta)$. In this field, these units make up a subgroup C_0 of $\mathbb{Z}[\zeta]^\times$ generated by C and ζ. Since $\zeta = \zeta^{dq}$, where d is the inverse modulo p of q, we now have that all units in C_0 are qth powers modulo q^2.

In particular, so is the unit $1+\zeta^q = \frac{1-\zeta^{2q}}{1-\zeta^q}$. This gives us a congruence of the form $1+\zeta^q \equiv \eta^q \pmod{q^2}$. A well-known property of binomial coefficients then implies that $(1+\zeta)^q \equiv \eta^q \pmod{q}$. By means of the Rule of Lifting the Exponent we therefore obtain

$$(1+\zeta)^q \equiv 1+\zeta^q \pmod{q^2}.$$

Hence the polynomial

$$f(T) = \frac{1}{q}((1+T)^q - 1 - T^q) \in \mathbb{Z}[T]$$

has ζ as a zero modulo q, and along with ζ also its conjugates ζ^k, $k = 1, \ldots, p-1$. Consider $f(T)$ as a polynomial over the field $\mathbb{Z}[\zeta]/Q$, where Q is a prime ideal factor of $\langle q \rangle$. Since this polynomial has $p-1$ distinct zeros, its degree $q-1$ is at least $p-1$. The primes p and q were assumed different, so that we must have $q > p$. But p and q can be interchanged, as is seen from (4.1). This shows that the above inequality cannot be true.

12. Concluding remarks

It is natural to ask whether Catalan's equation (1.1) has solutions in domains other than $\mathbb{Z}$. This question is briefly discussed by Ribenboim ([RIB], Appendix 1).

An immediate analog would be the integers in an algebraic number field F. As stated in [RIB], there is an extension of Tijdeman's result in this case. Indeed, under some mild conditions the solutions of

$$x^u - y^v = 1 \qquad (u > 1,\, v > 1), \tag{12.1}$$

where x and y are integers of F, can be bounded from above by an effective constant. Bounding x and y means here bounding the absolute values of all of their conjugates.

A problem worth studying might be the equation (12.1) with $u = p$, an odd prime, in the field $F = \mathbb{Q}(\zeta)$, where ζ is a primitive pth root of 1 as above. Since Thaine's theorem applies to this case, it may be possible to develop ideas similar to Mihăilescu's – perhaps starting even with a more general equation, like $x^p - y^v = \zeta$. However, the proof of Mihăilescu's theorem (Section 9) is very sensitive to all kinds of modifications, so that no straightforward generalization seems possible.

One may also wonder whether (12.1) can be solved in function fields. However, in the light of the following result by M.B. Nathanson [NAT] this problem is not very interesting. Let F be a field of characteristic not dividing v. If $u > 2$ and $v > 2$, the equation (12.1) with $x, y \in F(X)$ implies that x and y must be constants.

Acknowledgments

At early stages of this work, when examining the newest developments around Catalan's problem, I got valuable information from many colleagues. Special thanks go to Yuri Bilu, Radan Kučera, Preda Mihăilescu and René Schoof.

References

For a more complete bibliography see the references given in [RIB], [MI2], [BIL] and [M2].

[BIL] Y.F. Bilu, *Catalan's conjecture* [*after Mihăilescu*], Sém. Bourbaki, **55**ème année, n^o 909 (2002/03), 24 pp.

[BH] Y. Bugeaud, G. Hanrot, *Un nouveau critère pour l'équation de Catalan*, Mathematika **47** (2000), 63–73. MR **2003h:**11039

[CAS] J.W.S. Cassels, *On the equation* $a^x - b^y = 1$, *II*, Proc. Cambridge Philos. Soc. **56** (1960), 97–103. MR **22:**5610

[CAT] E. Catalan, *Note extraite d'une lettre adressée à l'éditeur*, J. Reine Angew. Math. **27** (1844), 192.

[CHE] E.Z. Chein, *A note on the equation* $x^2 = y^q + 1$, Proc. Amer. Math. Soc. **56** (1976), 83–84. MR **53:**7937

[COH] P.M. Cohn, *Algebra, Vol. 2*, John Wiley & Sons, Chichester – New York, 1977. MR **58:**26625

[DIL] K. Dilcher, *Fermat numbers, Wieferich and Wilson primes: computations and generalizations*, Public-key cryptography and computational number theory (Warsaw, 2000), de Gruyter, Berlin, 2001, pp. 29–48. MR **2002j:**11004

[GRE] R. Greenberg, *On p-adic L-functions and cyclotomic fields, II*, Nagoya Math. J. **67** (1977), 139–158. MR **56:**2964

[HY1] S. Hyyrö, *Über das Catalansche Problem*, Ann. Univ. Turku, Ser. A I no. 79 (1964), 8 pp. MR **31:**3378

[HY2] S. Hyyrö, *Über die Gleichung* $ax^n - by^n = z$ *und das Catalansche Problem*, Ann. Acad. Sci. Fenn., Ser. A I no. 355 (1964), 50 pp. MR **34:**5750

[IN1] K. Inkeri, *On Catalan's problem*, Acta Arith. **9** (1964), 285–290. MR **29:**5780

[IN2] K. Inkeri, *On Catalan's conjecture*, J. Number Theory **34** (1990), 142–152. MR **91e:**11030

[KER] W. Keller, J. Richstein, *Solutions of the congruence* $a^{p-1} \equiv 1 \pmod{p^r}$, Math. Comput. (to appear).

[KNR] J. Knauer, J. Richstein, *The continuing search for Wieferich primes*, preprint (2003).

[KO] Chao Ko [Ko Chao], *On the Diophantine equation* $x^2 = y^n + 1$, $xy \neq 0$, Sci. Sinica (Notes) **14** (1964), 457–460. MR **32:**1164

[LEB] V.A. Lebesgue, *Sur l'impossibilité en nombres entiers de l'équation* $x^m = y^2 + 1$, Nouv. Ann. Math. **9** (1850), 178–181.

[LET] G. Lettl, *A note on Thaine's circular units*, J. Number Theory **35** (1990), 224–226. MR **91h:**11118

[MW] B. Mazur, A. Wiles, *Class fields of abelian extensions of* $\mathbb{Q}$, Invent. Math. **76** (1984), 179–330. MR **85m:**11069

[MI1] M. Mignotte, *A criterion on Catalan's equation*, J. Number Theory **52** (1995), 280–283. MR **96b:**11042

[MI2] M. Mignotte, *Catalan's equation just before* 2000, Number Theory (Turku, 1999), de Gruyter, Berlin, 2001, pp. 247–254. MR **2002g:**11034

[M1] P. Mihăilescu, *A class number free criterion for Catalan's conjecture*, J. Number Theory **99** (2003), 225–231.

[M2] P. Mihăilescu, *Primary cyclotomic units and a proof of Catalan's conjecture*, preprint (September 2, 2002), submitted.

[NAG] T. Nagell, *Sur l'impossibilité de l'équation indéterminée* $z^p + 1 = y^2$, Norsk Mat. Forenings Skrifter, Ser. I no. 4 (1921), 10 pp.

[NAT] M.B. Nathanson, *Catalan's equation in* $K(t)$, Amer. Math. Monthly **81** (1974), 371–373. MR **49:**218

[RIB] P. Ribenboim, *Catalan's Conjecture*, Academic Press, Boston, 1994. MR **95a:**11029

[SCH] W. Schwarz, *A note on Catalan's equation*, Acta Arith. **72** (1995), 277–279. MR **96f:**11048

[STE] R. Steiner, *Class number bounds and Catalan's equation*, Math. Comput. **67** (1998), 1317–1322. MR **98j:**11021

[THA] F. Thaine, *On the ideal class groups of real abelian number fields*, Ann. of Math. **128** (1988), 1–18. MR **89m:**11099

[TIJ] R. Tijdeman, *On the equation of Catalan*, Acta Arith. **29** (1976), 197–209. MR **53:**7941

[WAS] L.C. Washington, *Introduction to Cyclotomic Fields*, 2nd ed., Springer-Verlag, New York–Berlin–Heidelberg, 1997. MR **97h:**11130

Department of Mathematics, University of Turku, FIN-20014 Turku, Finland
E-mail address: taumets@utu.fi