Squares mod p and the Golden Theorem

Kenneth A. Ribet
University of California, Berkeley, USA

The following article is based on a talk given by Professor Ribet at Victoria Junior College, Singapore, during the International Conference on Fundamental Sciences: Mathematics and Theoretical Physics, held from 13 to 17 March 2000 in Singapore. The conference was jointly organised by the Faculty of Science, National University of Singapore and the Isaac Newton Institute for Mathematical Sciences at Cambridge. Part of the programme of the conference was a series of talks at schools by some of the invited speakers, the talk by Professor Ribet was one of them.

Squares mod p and the Golden Theorem

Introduction

Because I've written several expository articles about Fermat's Last Theorem, I get lots of mail (and e-mail) from amateur mathematicians.

Today I'll discuss some questions that have been posed to me since my name was first linked with Fermat's Last Theorem.

Question 1

" You've solved $x^n + y^n = z^n$. What about $x^n + y^n = 2z^n$? "

Through reading and discussions with colleagues, I learned that the history of $x^n + y^n = 2z^n$ touches many of the same mathematicians who worked on Fermat's equation. We can paraphrase this equation as the statement that z^n is the *average* of x^n and y^n, or equivalently that the three nth powers lie in an arithmetic progression.

There are solutions with $xyz = 0$ and also those for which $x^n = y^n = z^n$. Are there other solutions?

For *squares*, it is easy to find non-trivial solutions. For example, 1, 25 and 49 lie in an arithmetic progression. However, Fermat proved that four distinct perfect squares cannot lie in an arithmetic progression. He proved also that three distinct *fourth powers* cannot lie in an arithmetic progression.

Cubes were treated by Legendre and Euler; nth powers with $n \leq 31$ were treated by Dénes in the 1950s. There are no non-trivial solutions when $3 \leq n \leq 31$.

In view of the results of Fermat and Legendre, to prove that there are no non-trivial solutions for $n > 2$, it's enough to study the case where n is prime. By adopting the ideas used to treat Fermat's equation and some techniques of Henri Darmon, I proved in 1994 that there are no non-trivial solutions when n is a prime that is congruent to 1 modulo 4.

In 1997, Darmon and Loïc Merel treated the remaining case when n is a prime that is congruent to 3 mod 4. ("Winding quotients and some variants of Fermat's last theorem.")

Question 2

" If you expand out $(x+y)^n$, the coefficient of $x^i y^{n-i}$ is the *binomial coefficient* $\binom{n}{i}$. For example, since $(x+y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5$, the binomial coefficients $\binom{5}{i}$ are 1, 5, 10, 10, 5, 1. I've been calculating these numbers on my laptop.

" I take a prime p congruent to 1 mod 4 and compute $b_p := \binom{\frac{p-1}{2}}{\frac{p-1}{4}}$:

p	5	13	17	29	37	$\cdots$
b_p	2	20	70	3432	48620	$\cdots$

These numbers are very big, so I reduce them modulo p.

" Let c_p be the residue of b_p mod p, represented as a negative number, if necessary, so that it's even. It seems to me that c_p is—up to sign—twice the odd number x that you get when you write $p = x^2 + y^2$ with x odd and y even:

p	5	13	17	29	37	41	53	$\cdots$
c_p	2	-6	2	10	2	10	-14	$\cdots$
x	1	3	1	5	1	5	7	$\cdots$

Why is this so? And what is the meaning of the minus signs? "

The relation between c_p and x was known in the 19th century. Both Cauchy and Gauss had proofs of the identity that my correspondent had uncovered. More precisely, Gauss proved a formula that yields the identity after a little bit of calculation; Cauchy has the identity pretty much as I've presented it.

Fermat proved that each p that is 1 mod 4 may be written in the form $x^2 + y^2$, but he gave no formula for x (or y). Cauchy viewed his identity as a formula for x.

To show you how subtle this business is, let me explain the recipe for the sign of c_p (the binomial coefficient that is reduced mod p and taken to be even). First, write $p = x^2 + y^2$ and think of x and y as determined only up to sign. Next, fix the sign of x temporarily so that $x + y - 1$ is divisible by 4.

For example, if $p = 53$, $x = \pm 7$, $y = \pm 2$, so we take $x = +7$. If $p = 41$, $x = \pm 5$, $y = \pm 4$, and we again take the "+." If $p = 37$, $x = \pm 1$, $y = \pm 6$, so we take $x = -1$ for the moment.

Finally, we change the preliminary sign of x if p is 5 mod 8 (but leave it alone if p is 1 mod 8). We change it for $p = 5$, $p = 13$, $p = 29$, $p = 37$, ... and leave it alone for $p = 17$, $p = 41$, $p = 73$, ...

The result is that this "tweaked" value of x is one-half the c_p that we displayed before. Let's do an example: $p = 13$. Then $x = \pm 3$ and $y = \pm 2$ at the start, so we choose initially $x = +3$ to make $x + y - 1$ be a multiple of 4. We change the sign, getting $x = -3$ because 13 is 5 mod 8. Indeed, $c_p = -6$.

Question 3

The next question that I'll discuss was posed by James Lawler, a 17-year-old student at Yale University. Lawler has been studying numbers mod p, where p is a prime number. For his question, we'll take $p > 3$: $p = 5, 7, 13, \ldots$. The numbers mod p can be added, subtracted and multiplied in the usual manner. You just toss away multiples of p, so that you end up with numbers between 0 and $p - 1$.

Lawler calculated the *squares* mod p. For example, when $p = 7$, the squares are 0, 1, 2 and 4.

The squares mod 11 are 0, 1, 3, 4, 5 and 9. Mod 13, the squares are 0, 1, 3, 4, 9, 10 and 12. The squares are 0, 1, 2, 4, 8, 9, 13, 15 and 16 modulo 17.

Number theory texts prove:

- Aside from 0, there are $\frac{p-1}{2}$ squares mod p;
- -1 mod p is a square if and only if p is 1 mod 4.
- The product of two squares mod p is again a square mod p.

A famous problem of the past centuries was to characterize those p for which a given number is a square. For example, when is 2 a square mod p? When is 17 a square mod p? (And so on.) Gauss was the first mathematician to give a complete solution to the problem.

Lawler has been adding up the squares mod p, tabulating the answer as an ordinary integer (not reducing mod p). For example, for $p = 11$, he computes $0 + 1 + 3 + 4 + 5 + 9 = 22$. For $p = 17$, he gets $0 + 1 + 2 + 4 + 8 + 9 + 13 + 15 + 16 = 68$.

It's not terribly hard to see that Lawler's sum L is always divisible by p. Lawler's question concerns the complementary factor:

p	5	7	11	13	17	19
L	$1 \cdot 5$	$1 \cdot 7$	$2 \cdot 11$	$3 \cdot 13$	$4 \cdot 17$	$4 \cdot 19$

Lawler noticed, first of all, that $L = p \cdot \frac{p-1}{4}$ when p is 1 mod 4.

Why is this true?

The point is that -1 is a square mod p in this case. Since products of squares are squares, the negatives of squares are squares as well. If x is a square between 1 and $p-1$, its negative is $p-x$. We can divide up the $(p-1)/2$ different non-zero squares into $(p-1)/4$ pairs, each pair summing to p. For example, with $p = 17$ we can re-write $0+1+2+4+8+9+13+15+16$ as $(1+16)+(2+15)+(4+13)+(8+9) = 4 \cdot 17$.

The remaining case where $p = 3$ mod 4 is harder because there is no obvious symmetry to exploit:

p	7	11	19	$\cdots$
L	$1 \cdot 7$	$2 \cdot 11$	$4 \cdot 19$	$\cdots$

Lawler initially thought that the complementary factor L was

$$\frac{p-3}{4},$$

as suggested by this table. I was skeptical, since I had never encountered this simple statement in a textbook.

Let's continue the table, tabulating the complementary factor L/p:

p	7	11	19	23	31	43	47	59	67	$\cdots$
L/p	1	2	4	4	6	10	9	13	16	$\cdots$
$\frac{p-3}{4}$	1	2	4	5	7	10	11	14	16	$\cdots$

It appears that L/p is always less than or equal to $(p-3)/4$ and that the identity $L/p = (p-3)/4$ is true fairly often. However, it is not true always!

In fact, it is known that $L/p = \dfrac{p-1-2h}{4}$, where h is an odd positive integer that depends on p. This h is a *class number* that measures the extent to which unique factorization fails for the numbers $n + m\frac{-1+\sqrt{-p}}{2}$, where n and m are integers. We have $h = 1$ for the first few values of p; $h = 3$ for $p = 23$ and $p = 31$; $h = 5$ for $p = 47$, and so on. (For a proof of the correct formula, see, e.g., Hermann Weyl's book *Algebraic theory of numbers.*)

Gauss studied the class number h and decided that $h \to \infty$ as $p \to \infty$. He suspected that $h = 1$ precisely for the following values of p: 3, 7, 11, 19, 43, 67, 163. When I was Lawler's age, it was known

only that there was at most one further value of p for which $h = 1$. However, when I was 18 or 19 years old, A. Baker and H. Stark proved independently that Gauss's list is in fact complete!

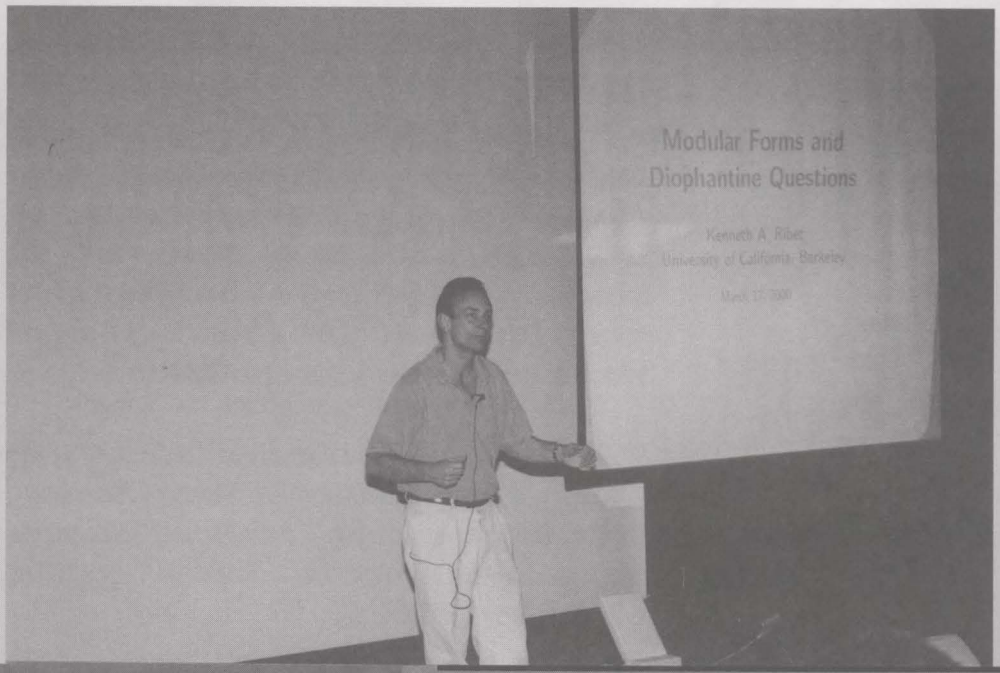
One Last Question

" I have a short proof of Fermat's Last Theorem. Would you like to see it? "

Answer: A recent book by Paulo Ribenboim, *Fermat's Last Theorem for Amateurs*, catalogs elementary methods that have been used to study Fermat's Last Theorem. All these methods seem to have limitations.

No one has found an elementary proof of FLT, and I would be extremely surprised to see one.

Squares mod p and the Golden Theorem



Professor Ribet is Professor of Mathematics in University of California, Berkeley, USA. He studies the theory of numbers and is most famous for his work related to Fermat's Last Theorem (FLT). Surely one of the most fascinating results in the history of mathematics, FLT states that the sum of the nth power of two whole numbers can never be a third nth power if n is a whole number greater than 2. In contrast, for n=2, there are infinitely many squares which are sums of two other squares, for example, $5^2 = 3^2 + 4^2$, $13^2 = 5^2 + 12^2$. FLT was first stated without proof by Fermat, an amateur French mathematician, in the 1 7th Century and had baffled mathematicians for more than three centuries before it was finally nailed down in 1994 by Andrew Wiles of Princeton University. In steep contrast to the simplicity of its statement, the resolution of Fermat's Last Theorem called for some of the most sophisticated mathematics ever created and is certainly one of the great achievements of 20th Century mathematics. The proof makes use of knowledge of the symmetry of certain objects in algebraic geometry known as elliptic curves. Professor Ribet's celebrated 1986 result provided an important step of the proof by showing the absence of such symmetries if FLT were not true.