Andrew Wiles and Fermat's Last Theorem

Ton Yeh

A great figure in modern mathematics

Sir Andrew Wiles is renowned throughout the world for having cracked the problem first posed by Fermat in 1637: the non-existence of positive integer solutions to $a^n + b^n = c^n$ for $n \geq 3$. Wiles' famous journey towards his proof – the surprising connection with elliptic curves, the apparent success, the discovery of an error, the despondency, and then the flash of inspiration which saved the day – is known to layman and professional mathematician alike. Therefore, the fact that Sir Andrew spent his undergraduate years at Merton College is a great honour for us.

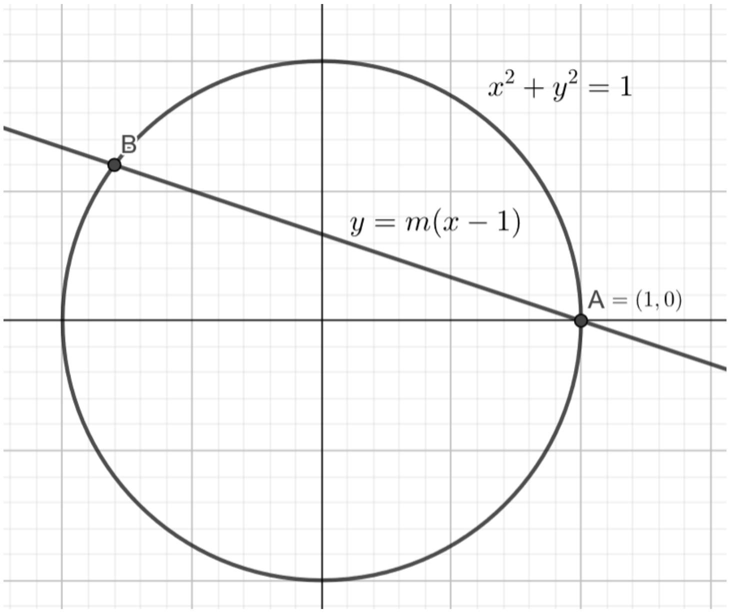
In this article we discuss simple number-theoretic ideas relating to Fermat's Last Theorem, which may be of interest to current or prospective students, or indeed to members of the public.

Easier aspects of Fermat's Last Theorem

It goes without saying that the non-expert will have a tough time getting to grips with Andrew Wiles' proof. What follows, therefore, is a sketch of much simpler and indeed more classical ideas related to Fermat's Last Theorem. As a regrettable consequence, things like algebraic topology, elliptic curves, modular functions and so on have to be omitted. As compensation, the student or keen amateur will gain accessible insight from this section; he will also appreciate the shortcomings of these classical approaches and be motivated to prepare himself for more advanced perspectives.

Our aim is to determine what can be said about the solutions to $a^n + b^n = c^n$ for positive integers a, b, c. Two immediate observations are in order. First, we can ignore $n = 1$. Second,an equivalent problem is to find the set of rational solutions to $x^n + y^n = 1$. Incidentally, the passage to rational numbers eliminates any significance of 'trivial' solutions (ak, bk, ck) corresponding to each known solution (a, b, c).

It behooves us to look at small exponents first. Therefore, try $n = 2$. It is well-known, even to children in primary schools, that $3^2 + 4^2 = 5^2$, and also perhaps that $5^2 + 12^2 = 13^2$. An extensive enough search will reveal hundreds of other solutions. But how many are there? A thousand? Ten billion? Infinitely many?



A slick proof ensures the correctness of the last answer. Look at the diagram above. Through point A at $x = 1, y = 0$ the line having rational slope m intersects the unit circle at a second point, called B. We want to show that B is a rational point, i.e. that both its coordinate are rational. But its x-coordinate satisfies

$$1 - x^2 = m^2(x-1)^2$$

and since $x = 1$ gives it one rational solution, the other solution has to be rational (since all the coefficients involved in the quadratic equation are rational). If $m \neq 0$, the y-coordinate satisfies

$$1 - y^2 = \left(\frac{y}{m} + 1\right)^2,$$

to which both solutions are again rational.

(If we explicitly compute the rational points on the circle, we will find a parametrisation for the original triple of integers (a, b, c); namely, $a = 2uv, b = u^2 - v^2, c = u^2 + v^2$, assuming u, v to be coprime. Here is an interesting consequence. Suppose c to be a square number, say $c = f^2$. Could either a or b be square? If a is square, then $u^2 + v^2 = f^2$ implying that $a = 2uv = 4wz(w^2 - z^2)$ for some coprime w, z. Now $w, z, (w-z), (w+z)$ are all coprime. They are all squares, since a is a square. Thus $w^2 - z^2 = w_0^4 - z_0^4$ is both the difference of two fourth powers and a square; also $w_0 < f$. Now suppose instead that b is square. Then $u^4 - v^4 = (u^2 - v^2)(u^2 + v^2)$ is a square, with $u < f$. So, starting with a solution to $f^4 - g^4 = h^2$ in the positive integers, for coprime f, g, h, we end up with another solution f', g', h' with $f' < f$. By impossible descent, no such f, g, h exists. This kills off $a^4 + b^4 = c^4$ and implies that Fermat's Last Theorem need only be checked for odd exponents.)

The same method of line-intersection applies to any non-trivial rational conic form that has at least one rational point. That is, if for some rational second-degree polynomial in two variables p, the set of points (x, y) with $p(x, y) = 0$ is more than just empty or a single point, and if that set has one rational point at all, it has infinitely many. Of course, it is not guaranteed that the set has a rational point at all. Consider $x^2 + y^2 = 3$, amounting to $a^2 + b^2 = 3c^2$ for pairwise coprime (a, b, c). The values on both sides, modulo 3, yield a contradiction right away.

A naive application of this method to $x^3 + y^3 = 1$, however, comes up against an obstacle. The obvious rational point to choose is either $(0, 1)$ or $(1, 0)$. In general, a line of rational slope through one of these meets the curve at two other points, and of course it cannot be deduced that either is rational. We could surmount the obstacle if the line going through both $(1, 0)$ and $(0, 1)$ hit a third point on the curve – which it does not – or if the tangent to the curve through $(1, 0)$ or $(0, 1)$ hit a second point – again it does not.

So other ideas are necessary – ideas which, it should be warned, go well beyond the usual high-school syllabus.

Just as the complex numbers reveal much about the real numbers and real functions, so perhaps statements about integers are illuminated by larger structures. To that end, it is easiestto go via number-fields. Recall that by definition, an algebraic number α is a complex root of some rational polynomial. It is often useful to look at fields made from adding to $\mathbb{Q}$ some extra irrational α. How is this constructed? Well, the rational polynomial p of smallest degree that has α as a root must be irreducible: in other words, it is not the product of two rational polynomials of strictly lower degree. Now, denoting the multiples of p as $\langle p(x)\rangle$, the sets $d(x) + \langle p(x)\rangle$ partition the rational polynomials. Check that naive addition and multiplication is well-defined, and then check that the resulting structure, written as $\mathbb{Q}[x]/\langle p(x)\rangle$, is in fact a field. It is not hard to prove that associating $x + \langle p(x)\rangle$ with α gives an isomorphism between $\mathbb{Q}[x]/\langle p(x)\rangle$ and $\mathbb{Q}(\alpha)$.

And just as we pass between rationals and integers, so we pass between the field of algebraic numbers and the ring of algebraic integers. A complex number β is an algebraic integer iff it is the root of some integer polynomial with leading coefficient 1, i.e. of the form $x^n + a_{n-1}x^{n-1} + ... + a_0$ with all $a_i \in \mathbb{Z}$. (Yes, the algebraic numbers are a ring. The best explanation involves finitely-generated $\mathbb{Z}$-modules: see any good textbook for details.) Much of algebraic number theory studies the subring of algebraic integers contained in $K = \mathbb{Q}(\alpha)$, called $\mathbb{O}_K$, and we hope to examine its structure.

The choice $\alpha = e^{2\pi i/p}$, for odd prime p, seems suitable for Fermat's Last Theorem. Following convention, we denote this p-th root of unity by ξ. Indeed, $\mathbb{O}_{\mathbb{Q}(\xi)} = \mathbb{Z}[\xi]$, a technical result not worth pursuing. The apparent value of this choice lies in the factorisation

$$a^p + b^p = (a + b)(a + b\xi)(a + b\xi^2)...(a + b\xi^{p-1}),$$

which one hopes to compare fruitfully to c^p, by associating the p-th power of each factor of c

with one of the factors $(a + b\xi^i)$. It is desired that the $(a + b\xi^i)$ have no common factors and therefore should be p-th powers, and that some contradiction should be obtained. Thus ran Lamé's claimed proof of Fermat's Last Theorem in his 1847 address to the Paris Academy, which attracted much attention at the time.

Unfortunately, as Liouville pointed out to him soon afterwards, a naive application of this method assumes that $\mathbb{Z}[\xi]$ is a unique factorisation domain or UFD. To see what this means, $\mathbb{Z}$ is a UFD: each non-zero integer is an almost unique finite product of primes (non-invertible elements p such that $p|ab \implies p|a$ or $p|b$). We can permute the primes, we can multiply some of them by invertible elements (here, only 1 or -1), but otherwise the product is unique. This looks like a natural condition to impose, but consider $\mathbb{Z}[\sqrt{-5}]$, wherein

$$(1 + \sqrt{-5})(1 - \sqrt{-5}) = (2)(3) = 6. \tag{1}$$

None of the four factors involved is prime, and any other factorisation of 6 is only trivially different. Importantly, this allows $(2 + 2\sqrt{-5})(3 - 3\sqrt{-5}) = 6^2$ even though neither factor on the left is a square.

The sad truth is that $\mathbb{Z}[\xi]$ can be troublesome in the same way: Kummer discovered in the 1840s that it is not a UFD for $p = 23$.

Kummer did not stop there, though; his efforts to make progress on Fermat's theorem, despite the disappointing discovery, guided him towards a sturdier notion of unique factorisation for the context of number-rings – concerning the ideals of $\mathbb{O}_K$.An ideal I of a ring R is a subset of R such that $I + I \subseteq I$ and $aI \subseteq I$ for all $a \in R$. The product of two ideals I and J is the set of all finite sums of $i_n j_n$ with $i_n \in I, j_n \in J$. The 'correct' type of factor, this time, is the maximal ideal: an ideal I such that $I \neq 0$, $I \neq R$, and for any ideal J, we have $I \subseteq J \implies J = R$. A triumph of algebraic number theory is that each non-zero ideal in a number-ring is a unique product of maximal ideals. This theorem yields a nice interpretation of non-unique factors of elements, like equation (1) in $R = \mathbb{Z}[\sqrt{-5}]$. If we denote by $\langle a_1, ..., a_n\rangle$ the smallest ideal containing $a_1, .., a_n$, then of course $\langle 1 + \sqrt{-5}\rangle\langle 1 - \sqrt{-5}\rangle = \langle 2\rangle\langle 3\rangle$. But in fact, as the reader can verify, each side expresses the product of four maximal ideals

$$\langle 2, 1 + \sqrt{-5}\rangle^2 \langle 3, 1 + \sqrt{-5}\rangle\langle 3, 1 - \sqrt{-5}\rangle$$

in two different ways.

It is also worth noting that the quotient of the non-zero ideals by the principal ideals – that is, the classes arising from the equivalence relation $r(I, J) \iff \exists \alpha, \beta \in \mathbb{O}_K - \{0\} : \alpha I = \beta J$ and multiplication defined in the usual way – is a finite abelian group, known as the class group. Numerous techniques have been developed to ease the computation of the size of the class group, called the class number. This number is of great importance. If we know it to be equal to 1, as is the case for $\mathbb{Z}[\xi]$ for $p = 3, 5, 7, 11$ and more, then every ideal in the number-ring is equivalent to the ring; such a number-ring is described as a principal ideal domain or PID, and every PID is a UFD. But Kummer also discovered, through an ingenious argument that turns the factors in $(a+b)(a+b\xi)(a+b\xi^2)...(a+b\xi^{p-1})$ into the relevant ideals, that if p does not divide the class number of $\mathbb{Z}[\xi]$, then $a^p + b^p = c^p$ has no solutions in the positive integers.

Conveniently, most of the primes below 100 count among these 'regular primes'. The fly in the ointment is that some primes do divide the class number of $\mathbb{Z}[\xi]$! 37 is a famous example. But I hope that this section has shown you the fascinating side of the classical ideas related to Fermat's Last Theorem, and (indirectly) a sense that if even these do not suffice to tackle the theorem, then how much deeper the modern number theorists had to go!