Fermat's last theorem, elliptic curves and modular forms

Ashwin Iyengar

Contents

1 Introduction

You may have heard of the following theorem.

Theorem 1.0.1. *If $n > 2$ is an integer and a, b, c are three positive integers, then $a^n + b^n \neq c^n$.*

Before thinking about how to prove this, let's examine it more closely.

- If $n = 1$, the (infinitely many) solutions are easy to classify.
- If $n = 2$ the solutions are called *Pythagorean triples*, and it is known completely how to classify them. There are a few different (essentially similar) proofs, but the basic idea is to factor the equation as

  $$a^2 + b^2 = (a + ib)(a - ib) = c^2$$

  (where $i^2 = -1$) and use properties of "prime numbers" in the ring

  $$\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\},$$

  which is the ring of *Gaussian integers*. An elementary formula is

  $$m > n > 0 \text{ with } m, n \text{ coprime (and one of them even) gives } a = m^n - n^2 \text{ and } b = 2mn$$

Let's focus more on the $n > 2$ case. First let's perform a reduction step. If $n = pm$ for p an odd prime, then $x^n + y^n = z^n$ implies $(x^m)^p + (y^m)^p = (z^m)^p$. If not, then $n = 2^r$, and then $x^n + y^n = z^n$ implies $(x^{2^{r-2}})^4 + (y^{2^{r-2}})^4 = (z^{2^{r-2}})^4$. So we are reduced to proving Fermat's Last Theorem for $n = 4$ and $n = p$ an odd prime.

Fermat famously claimed to have proven his "Last Theorem", but unfortunately his proof didn't fit into the margins. It is widely suspected that he had a proof which worked in certain cases, but which fails in general. The case $n = 4$, for which Fermat gave a complete proof, can be done by the method of *infinite descent* — the idea is that you can extract an infinite descending sequence of Pythagorean triples using the existence of a solution, which is a contradiction.

So this then leaves the case of $n = p$ an odd prime.

- Sophie Germain proved a number of cases using some ingenious tricks for certain primes of special forms — namely she showed that for primes less than 270 the theorem is true if $p \nmid xyz$, and she showed many other cases as well (possibly infinitely many, but this is a conjecture).
- A mathematician named Ernst Kummer managed to find a proof of this theorem when p is what's called a *regular prime*, following a strategy of Gabriel Lamé. I won't define regularity, but it has to do with the extent to which a ring like $\mathbb{Z}[\zeta_p]$ enjoys the property of **unique factorization**. The proof of this theorem then follows from fairly straightforward arguments that just involve some elementary number theory manipulations.

  But not all primes are regular! In fact it's an open question as to whether there are infinitely many of them, although it's conjectured to be the case. On the other hand, one does know that there are infinitely many **irregular** primes — the first one is 37. The elementary argument in the regular case completely fails in the non-regular case, so what to do?
- Gerd Faltings proved a conjecture of Mordell that says that there can only be at most finitely many solutions for a given n.
- Computers did some verification.

But despite many efforts, the theorem remained unproven for a very long time. It was only proven in the mid-90s by Andrew Wiles, assisted by his student Richard Taylor. Their methods were vastly different though, so let's forget about this equation for a moment, and try to understand some other math which will help us.

2 Detour: elliptic curves

Definition 2.0.1. An elliptic curve over a field K (of characteristic not equal to 2 or 3) is the solution set of the polynomial

$$y^2 = x^3 + ax^2 + b$$

where $4a^3 + 27b^2 \neq 0$. Alternatively, if you can factor the degree 3 polynomial on the right (this is a special case) you get

$$y^2 = (x-a)(x-b)(x-c)$$

and the condition $4a^3 + 27b^2 \neq 0$ translates to $a, b, c \in K$ are pairwise distinct elements.

Note that in this definition we imagine that the elliptic curve is sitting in the projective plane over the field K: this can be thought of as the space of lines in 3-space over K passing through the origin. The condition a, b, c being pairwise distinct is there to ensure that the curve is not *singular*.

Example 2.0.2. Let's plot the graphs of some elliptic curves over $\mathbb{R}$.

- If $y^2 = x^3$, you get a cusp.
- If $y^2 = x(x-1)^2$, the curve passes through itself.
- If $y^2 = x^3 + x^2 + 1$ you get something that looks like a single line
- If $y^2 = x(x+1)(x+2)$ you get a line and a loop.

Note that in each of these examples there's a "point at ∞" — to see this, you need to work in projective space, but never mind that for now.

Note that the particular picture you get from the equations above is a consequence of the fact that we're working over the reals. If we consider the solutions to the equation in the complex numbers, what happens? Well, instead of drawing a picture in $\mathbb{R}^2$ (plus some infinity boundary extra thing) we're working in $\mathbb{C}^2$. This has 4 real dimensions. The one complex equation has a real and imaginary part, so that's really 2 equations, so you should expect the elliptic curve to look like some kind of surface. This surface should be smooth because of the discriminant condition. So you can think of it as a *complex manifold* (Riemann surface).

It turns out (this is called *complex uniformization*) that every elliptic curve is isomorphic to a 2-torus as a complex manifold. If you remember more structure you can distinguish them.

It turns out the points on an elliptic curve form a group, and you can add the points to each other, giving some more rich geometric structure. But this will be irrelevant for us, so we'll ignore this for now.

3 Another detour: modular forms

Now let's study something in a (a priori) completely different direction.

Let $\mathcal{H}$ denote the complex upper half plane, and consider the group

$$\mathrm{SL}_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{Z} ad - bc = 1 \right\}.$$

This is a group under multiplication. This group acts on $\mathcal{H}$ as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau = \frac{a\tau + b}{c\tau + d}$$

To get a feel for this action, one can first show that $\mathrm{SL}_2(\mathbb{Z})$ is generated by two elements

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

So how do they act? We get

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \tau = \tau + 1$$
$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \tau = -\frac{1}{\tau}$$

Now, one can use this to show that if you start with any $\tau \in \mathcal{H}$, you can always act by these two generators to get it inside of a *fundamental domain* (draw the picture).

Definition 3.0.1. A modular form (of level 0 and weight $k \geq 0$) is a holomorphic function $f : \mathcal{H} \to \mathbb{C}$ satisfying

$$f(\gamma \cdot \tau) = (c\tau + d)^k f(\tau)$$

where $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$, and satisfying a growth condition.

Now note that a modular form is periodic, with a period of one. This means that it has a *Fourier expansion*. In other words, we can write

$$f(\tau) = \sum_{n=0}^{\infty} a_n e^{2\pi i n \tau} = \sum_{n=0}^{\infty} a_n q^n$$

where we usually abbreviate $q = e^{2\pi i n}$. So modular forms can be understood in terms of these q-expansion coefficients.

Theorem 3.0.2 (Hecke)**.** *For certain modular forms (called "newforms", which satisfy $a_0 = 0$ and $a_1 = 1$), the coefficients a_n are actually algebraic integers, meaning they solve polynomials in $\mathbb{Z}[x]$.*

This is an incredibly strong and interesting theorem! The point is that you started with some very complex analytic thing and it turned out to be algebraic, which is rather unexpected. There's a very good and systematic explanation for why this turns out to be true, which makes it less unexpected, but it's beyond the scope of this talk.

4 Miracles

Okay, here's a miracle that occurs.

Take the elliptic curve $E : y^2 + y = x^3 - x^2$. This doesn't look like what we wrote before, but you can make it into what we wrote before by a change of coordinates.

Now, let's count the number of solutions to E, but do it *mod* p for various prime numbers.

Example 4.0.1. Take $p = 2$. How many solutions are there? For any elliptic curve, there's always a point at ∞. Then the possibilities are $(0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$. Which satisfy the equation?

We can do the same for $p = 3$, $p = 5$ etc. When $p = 11$ we get a problem, which is related to the fact that the discriminant of this elliptic curve happens to be 11 (which means that the equation over the field $\mathbb{F}_{11}$ no longer defines a smooth curve), but let's just ignore this for now.

We can make a table of values.

p	$p + 1 - \#E(\mathbb{F}_p)$
2	-2
3	-1
5	1
7	-2
11	1
13	4
17	-2
19	0
23	-1
29	0
31	7
37	3
41	-8
43	-6
47	8
53	-6
59	5
61	12
67	-7
71	-3

Why did I do this? Well here's a specific modular form.

$$f(\tau) = q \prod_{n=1}^{\infty} (1 - q^n)^2 (1 - q^{11n})^2$$

If we expand this out, we get $f(\tau) = \sum_{n=0}^{\infty} a_n q^n$, and if we plot the a_p for p prime, we get the exact same table as above (you can check this on the LMFDB).

Coincidence? Let's make a definition.

Definition 4.0.2. An elliptic curve (with rational number coefficients) is **modular** if there is some modular form for which $a_p = p + 1 - \#E(\mathbb{F}_p)$. (I'm omitting some important technical details and hypotheses here, but this is meant to just be a sketch.

Theorem 4.0.3 (Wiles, Taylor–Wiles, Breuil–Conrad–Diamond–Taylor)**.** *Every elliptic curve over* $\mathbb{Q}$ *is modular.*

This theorem, which before was called the Taniyama–Shimura–Weil conjecture, is the culmination of a lot of extremely hard work that is way way outside the scope of this talk.

5 Back to Fermat

We have taken 2.5 detours, and still don't know why. Let's explain that now.

Recall that we want to show that if p is an odd prime and $a, b, c > 0$ satisfy $a^p + b^p = c^p$ then we get a contradiction. So suppose it is the case that $a^p + b^p = c^p$. Yves Hellegouarch asks us to consider the elliptic curve

$$y^2 = x(x - a^p)(x - c^p)$$

Note that this defines an elliptic curve over $\mathbb{Q}$, since the three roots are distinct. Gerhard Frey pointed out that this curve should have strange properties, relating to something called stability. Then Jean-Pierre Serre conjectured further that those properties should imply that the elliptic curve is not modular — this became known as the ϵ-conjecture and Serre made some progress towards a proof. Finally, Ken Ribet completed the proof in the late 1980s.

Of course, this happened before Wiles and Taylor–Wiles's proof, so that provided the final missing ingredient.

So we now know the (somewhat useless) fact that Fermat's last theorem is true.