Faltings' Theorem, or, How Geometry Makes Everything Better

S. M.-C.

12 March 2016

Abstract

An important theme in number theory is the surprising and powerful applications of geometry. I will talk about how polynomials can be thought of geometrically, and how this can be used to understand their solutions; especially Faltings' theorem, which tells us (very roughly) how many rational number solutions a polynomial has based on how many "holes" it has geometrically. (This talk won't assume any knowledge beyone polynomials and rational numbers; knowledge of complex numbers will be helpful).

Polynomials capture the basic arithmetic operations of addition and multiplication, and many questions about how integers or rational numbers behave can be thought of as questions about solving polynomial equations (in integers or rational numbers). One of the oldest such questions is producing Pythagorean triples. Recall that the three sides of a right triangle are related by the equation $x^2 + y^2 = z^2$. We can ask: what are the right triangles with integer side lengths? Or: which squares can be written as a sum of two squares? Or: what are the integer solutions of $x^2 + y^2 = z^2$?

1 Pythagorean Triples and the case of Degree 2

Here is the solution. Let m, n be integers (with $m > n$). Then $m^2 - n^2,\ 2mn,\ m^2 + n^2$ form a Pythagorean triple, and every Pythagorean triple is one of these (or a multiple of one of these). It's easy to check that this is actually a Pythagorean triple:

$$(m^2 - n^2)^2 + (2mn)^2 = m^4 - 2m^2n^2 + n^4 + 4m^2n^2 = m^4 + 2m^2n^2 + n^4 = (m^2 + n^2)^2.$$

It's a bit harder to check that every primitive Pythagorean triple arises in this way. (By "primitive" I mean that the triple of numbers a, b, c have no common factor). If a, b, c is a primitive Pythaogrean triple, then we can write

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

Now set

$$\frac{m}{n} = \frac{c + a}{b} = \frac{b}{c - a}.$$

Solving

$$\frac{c}{b} + \frac{a}{b} = \frac{m}{n}, \qquad \frac{c}{b} - \frac{a}{b} = \frac{n}{m}$$

gives

$$\frac{c}{b} = \frac{m^2 + n^2}{2mn}, \qquad \frac{a}{b} = \frac{m^2 - n^2}{2mn}.$$

We can do a bunch of modular arithmetic to check that all these fractions are reduced, so we can equate the numerators and denominators to get

$$a = m^2 - n^2, \quad b = 2mn, \quad c = m^2 + n^2.$$

But this is just a bunch of computations out of nowhere, that gives us no insight into what's going on. Is there a better way to see what's going on? And for that matter, how could we come up with these formulas in the first place? If we were presented with a different equation, say $2x^2 + 3y^2 = 7z^2$, could we find all the integer solutions of that one also?

The way to see what's really going on is through geometry. This path might take a bit more effort to set up, but it's a much better way to think about this problem.

First of all, observe that integer solutions of $x^2 + y^2 = z^2$ are the same as rational solutions of $x^2 + y^2 = 1$, because if we have a rational solution

$$\frac{a^2}{c^2} + \frac{b^2}{d^2} = 1$$

then we can clear denominators to get an integer solution

$$(ad)^2 + (bc)^2 = (cd)^2;$$

conversely, if we have an integer solution $a^2 + b^2 = c^2$ we can get a rational solution

$$\frac{a^2}{c^2} + \frac{b^2}{c^2} = 1.$$

So instead of asking for integer solutions of $x^2 + y^2 = z^2$, we can rephrase our question as asking for rational solutions of $x^2 + y^2 = 1$.

Now: what is the geometric interpretation of a polynomial? It's probably familiar! If we have a polynomial in two variables x, y, we can plot its solutions in the x-y-plane to get a curve. In the case of $x^2 + y^2 = 1$, we get a circle. We can translate our question to geometry by asking for rational points on this circle (by "rational point" I mean a point with rational coordinates), since these are the same as rational solutions to the equation $x^2 + y^2 = 1$.

We can easily find all rational points on the circle with a bit of geometry. Imagine that we draw a line through the circle, intersecting it twice. I claim that if one of the intersections is a rational point, and the slope of the line is rational, then the other intersection is also rational.

To see this, first note that if a polynomial has rational coefficients, and all but one of its roots are rational, then the last root must be rational as well. This is because, for example, the second coefficient is the sum of the roots. To be more precise, in the case of a quadratic polynomial, we have

$$(x - \alpha)(x - \beta) = x^2 - (\alpha + \beta)x + \alpha\beta.$$

So if α and $\alpha + \beta$ are rational, then $\beta = (\alpha + \beta) - \alpha$ is the difference of two rational numbers, thus rational.

Now back to geometry. Say our line is $y = ax + b$ with a, b rational numbers. Then the two intersections of the line and circle are the two roots of

$$x^2 + (ax + b)^2 = 1,$$

the equation we get by substituting $y = ax + b$ into $x^2 + y^2 = 1$; a standard way to solve systems of equations. Expanding, we get a polynomial with rational coefficients. If one intersection is a rational point, i.e. this polynomial has one rational root, then the other is rational also.

This gives us a process for producing rational points. Namely, if we can find a single rational point, then every line through that point with rational slope will intersect the circle in another rational point. Furthermore, the line between two rational points evidently has rational slope, which shows that every rational point arises in this way.

Let's see how this works for our circle $x^2 + y^2 = 1$. For starters, we need a rational point. The point $(-1, 0)$ will do. Now we draw a line $y = \frac{n}{m}(x+1)$ through this point with rational slope. The other intersection of the line with our circle is the other root of

$$x^2 + \left(\frac{m}{n}(x+1)\right)^2 = 1.$$

If we expand this out and solve for the other root (besides $x = -1$), we find it is $x = \frac{m^2-n^2}{m^2+n^2}$, and plugging this in to $y = \frac{n}{m}(x+1)$ we find the corresponding y-value is $y = \frac{2mn}{m^2+n^2}$. Thus any integers m, n give us a rational point

$$\left(\frac{m^2-n^2}{m^2+n^2}, \frac{2mn}{m^2+n^2}\right),$$

which (by clearing denominators) corresponds to the Pythagorean triple

$$m^2 - n^2, \quad 2mn, \quad m^2 + n^2.$$

The geometry explains the mysterious answer to producing Pythagorean triples!

Notice also that nothing we did depended specifically on the circle; the essential thing is that the line we drew intersected our curve in exactly one more point, which only depends on the equation having degree 2. So if we have any polynomial of degree 2 in two variables, we can use the same method. If we can find a single rational point, then by drawing lines of rational slope through this point and intersecting them with our curve, we can produce all rational points on the curve, i.e. rational solutions to our polynomial.

2 Degree 3

Let's move on to degree 3 and see how much we can say. For example, consider the curve

$$y^2 + xy = x^3 - 2x + 1.$$

If we try to apply the same method, we'll see that it doesn't work. Starting from a rational point, we can draw a line (with rational slope); but it will generally intersect our curve in two more points, rather than one, and there's no guarantee that they will be rational.

But there is still something we can do: if we draw the tangent line to our curve at a rational point, it will intersect our curve in only one other point, and this is guaranteed to be a rational point. Alternatively, if we have two rational points, then we can draw the line connecting them, and this also will intersect our curve in only one more point, which will be another rational point. So we do have methods for producing rational points, but not as easily as in the case of degree 2.

In fact, it turns out that the curve above has infinitely many rational points: starting from the point $(1, -1)$, we can draw a tangent line to get the point $(0, 1)$; if we draw the tangent at this new point, we get the point $(\frac{3}{4}, -\frac{1}{8})$; the tangent at this point gives us $(-\frac{111}{64}, \frac{413}{512})$; etc.

On the other hand, the degree 3 curve

$$y^2 + y = x^3 - x^2$$

has only 4 rational points: $(0,0)$, $(0,-1)$, $(1,0)$, $(1,-1)$. If we draw any tangents, or lines connecting two of these points, we'll just keep getting the same four points back again.

We saw that every curve defined by a (rational) degree 2 polynomial has infinitely many rational points (if it has any at all), but in this case it's possible to have finitely many or infinitely many rational points. Can we say anything about how many rational points any given curve has? We can, and it depends on a more subtle geometric property.

3 Polynomials Geometrically

Up until now we've been looking at (and drawing) only the points on our curves with real coefficients. To understand what happens in general, we have to expand to using complex numbers. If you're familiar with complex numbers, great; if not, here's all you need to know for now. We get the complex numbers by taking the real numbers and adding a square root of -1, which we call i, and every complex number can be written as $a + bi$ with a, b real numbers. The point is that the complex numbers are two-dimensional; or to be precise, the complex numbers are two-real-dimensional, but only one-complex-dimensional.

Just like the real picture, we should think about the set of complex points of our curve as a geometric object. But now it will be one-dimensional in the complex sense, and two-dimensional in the real sense, forming what we think of as a surface.

There are many geometries that a surface can have, but they have one distinguishing feature more obvious than the rest: the number of holes, called the genus. Every surface is topologically (i.e. can be bent and squished into) a sort of baguette with some holes.

The first example we looked at, the curve defined by the equation $x^2 + y^2 = 1$, as a complex curve is a sphere; it has no holes, i.e. genus 0. The second two curves, $y^2 + xy = x^3 - 2x + 1$ and $y^2 + y = x^3 - x^2$, as complex curves both have one hole, genus 1, like a doughnut.

4 Faltings' Theorem

The examples we have seen are representative:

- if a curve has genus 0, then it has infinitely many rational points (or none at all);
- if a curve has genus 1, then it has finitely many or infinitely many rational points (or none at all).

The case of higher genus is Faltings' theorem:

Theorem 1 (Faltings)**.** *If a curve has genus 2 or more, then it has finitely many rational points (or none at all).*

So in some sense genus 0 and 1 are the exceptional cases; in all the rest, curves have finitely many rational points. This is a beautiful example of how geometry can be used to understand problems in number theory.

As an application, recall that Fermat's Last Theorem states that the equation $x^n + y^n = z^n$ has no integer solutions for $n \geq 3$. Faltings' theorem does not imply the whole of Fermat's Last Theorem, but it does imply that (for $n \geq 4$) that the equation $x^n + y^n = z^n$ has only finitely many integer solutions (up to multiples).