Pythagorean Triples and Fermat's Last Theorem

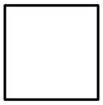
A. DeCelles
Notes for talk at the Science Speakers Series at Goshen College
Document created: 05/10/2011
Last updated: 06/05/2011 (graphics added)

It is an amazing fact of human experience that, to answer an innocent-sounding question, it may be necessary to develop significant new ideas. The field of number theory can be described as the mathematics that has developed in an effort to answer simple questions about the most elementary of mathematical objects: the whole numbers. In this talk I will trace a few such threads.

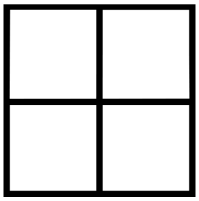
1. Squares, diagonals, and rationality

To set the stage for the discussion of Pythagorean triples (which are, as you may have guessed, related to the oh-so-familiar Pythagorean theorem), I will first discuss squares and their diagonals, in the style of the ancient Greeks (minimal mathematical notation, no variables.) (In fact, I am drawing from a conversation between Socrates and a slave boy in Plato's *Meno*.)

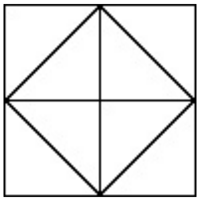
Consider a square of side length, say, two feet. Then the size of the square is four square feet.



To make a square of double the size, we might naively guess that we ought to double the side length (four feet), but this results in a square of size sixteen, which is twice as large as we would like.



But notice that if we cut each quadrant in half, along the diagonals drawn below,



the square in the middle is of size eight (half sixteen, twice four)! So to double a square, use the diagonal.

But perhaps this geometric characterization of the answer is not completely satisfactory. Perhaps we would like to know the *length* of the diagonal. Clearly it is between two and four. Perhaps it is three? This cannot be so, because the square of three is nine, which is *not* eight. So the length of the diagonal is not a whole number.

We might hope that if we simply consider a square with a larger side length, the diagonal may "come out even," like, for example, if you double a recipe so that you don't have to measure any half-cups. However, as the Pythagoreans showed, using the argument of *infinite descent* (a precursor to the method of mathematical induction), there is no way to make a square whose side length and whose diagonal are both whole numbers!

We explain the proof, now permitting ourselves to use a more modern style. We have shown that if we have a square of side length a and diagaonal of length b, then the square of b is double the square of a, i.e.

$$2a^2 \;=\; b^2$$

(You may recognize this as the simplest case of the Pythagorean theorem.) Suppose that a and b are whole numbers. Then b^2 is even, and (since evens square to evens and odds to odds) b is even as well. Say

$$b = 2b_1$$

Substituting in,

$$\begin{aligned} 2a^2 &= (2b_1)^2 \;=\; 4b_1^2 \\ a^2 &= 2b_1 \end{aligned}$$

By the same argument, a is even, say $a = 2a_1$, and

$$\begin{aligned} (2a_1)^2 &= 2b_1 \\ 2a_1^2 &= b_1^2 \end{aligned}$$

Thus b_1 is even, say $b_1 = 2b_2$, and

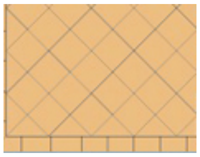
$$\begin{aligned} 2a_1^2 &= (2b_2)^2 \\ a_1^2 &= 2b_2^2 \end{aligned}$$

In this way we construct sequences of whole numbers $b > b_1 > b_2 > \ldots$ and $a > a_1 > a_2 > \ldots$. But we cannot have infinitely decreasing sequences of whole numbers!

As I mentioned above, this proof was known to the school of Pythagorus, and it was *deeply disturbing* to them, since their (very natural!) concept of number was *whole number* (and, in an extended sense, *ratios* of whole numbers.) This led to a rift between arithmetic and geometry. They developed a way to add and multiply lengths geometrically without quantification of lengths and areas. It wasn't until much later, after algebra had been developed, that Decartes invented analytic geometry (putting coordinates on geometric objects) and acheived a sort of reconcilliation between arithmetic and geometry.

In modern terms, the proof given above is a proof for the *irrationality* of $\sqrt{2}$. The ancient Greek rejection of irrational numbers is understandable because we now know that any real number can be approximated aribtrarily well by rationals. It wasn't until 1858 that Dedekind gave a precise description of the real numbers.

Exercise 1. I once used the tile floor in our school auditorium to find a rational approximation for $\sqrt{2}$. Can you see how? (You might have to use some graph paper to make a picture with more tiles.)



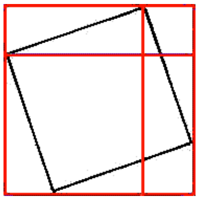
2. Pythagorean triples: Diophantus' chord method

Recall the Pythagorean theorem:

$$a^2 + b^2 \;=\; c^2$$

where a, b and c are the lengths of the sides of a right triangle, with c being the length of the hypotenuse.

Exercise 2. Can you see how the following picture proves the theorem?



We might ask if there are whole numbers a, b, and c satisfying this equation. (We have already shown that if $a = b$, this is impossible.) Well, sure: the triple three, four, five is one example. Notice that given one Pythagorean triple, multiplying through by a constants yields others. So, in fact, there are infinitely many Pythagorean triples. The ones that can't be obtained by multiplying another triple by a constant are called *primitive.*

Diophantus (c. 200 BC) had a way of generating Pythagorean triples by looking at rational points on circles. Dividing the equation

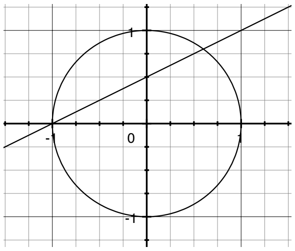
$$a^2 + b^2 \;=\; c^2$$

through by c^2 yields

$$\big(a/c\big)^2 + \big(b/c\big)^2 \;=\; 1$$

So solving the original equation in integers is equivalent to solving the equation $x^2 + y^2 \;=\; 1$ in *rationals.* But this is precisely the equation of a *circle* of radius one. Thus, finding Pythagorean triples is equivalent to finding *rational points* on the unit circle. (Algebraic geometry: finding rational points on curves!)

Claim 1. *Rational points on the unit circle are in one-to-one correspondence with lines of rational slope through the point* $(-1, 0)$.



Proof. Clearly if R is a rational point on the unit circle, the line passing through $(-1, 0)$ and R has rational slope. To prove the converse, suppose t is the rational slope of a line through $(-1, 0)$. Then, using point-slope form $((y - y_o) = m(x - x_o))$, the equation of the line is

$$y \;=\; t\,(x + 1)$$

How do we find the point at which this line intersects the unit circle? Well, an intersection point (x, y) must satisfy the equation of the line *and* the equation of the circle. So use substitution:

$$x^2 \;+\; \big(t(x+1)\big)^2 \;=\; 1$$

Solving for x:

$$x^2 \;+\; t^2(x^2 + 2x + 1) \;=\; 1$$
$$(1 + t^2)x^2 \;+\; (2t^2)x \;+\; (t^2 - 1) \;=\; 0$$

Using the quadratic equation, we get,

$$x \;=\; \frac{-2t^2 \pm \sqrt{4t^4 - 4(1+t^2)(t^2-1)}}{2(1+t^2)} \;=\; \frac{-2t^2 \pm \sqrt{4t^4 + 4(1-t^4)}}{2(1+t^2)} \;=\; \frac{-t^2 \pm 1}{1+t^2}$$

i.e. $x = -1$ or $(1-t^2)/(1+t^2)$. Plugging back in to find y,

$$y \;=\; t \cdot \left(\frac{1-t^2}{1+t^2} + 1 \right) \;=\; \ldots \;=\; \frac{2t}{1+t^2}$$

Thus the intersection point (x, y) has rational coordinates. □

In fact (more algebra), if we let $t = v/u$, for u, v whole numbers, then the rational point is

$$\left(\frac{u^2-v^2}{u^2+v^2}, \frac{2uv}{u^2+v^2} \right)$$

Multiplying by $w(u^2+v^2)$, where w is a whole number, we can see that this correpsonds to the integer point

$$\big((u^2-v^2)w,\, 2uvw \big)$$

on a circle of *integer* radius $r = w(u^2+v^2)$. If we omit the w, we get all the primitive Pythagorean triples. (Euclid had a parametrization like this one hundred years before Diophantus.)

A natural next question is whether we can "split a cube into two cubes," i.e. whether there are integer solutions to

$$x^3 + y^3 \;=\; z^3$$

or even more generally, whether there are integer solutions to

$$x^n + y^n \;=\; z^n \qquad (n > 2)$$

While reading Diophantus, Fermat made a now-famous margin note, claiming to have a marvelous proof (too long to fit in the margin) that there are no such solutions. While Euler and Lagrange proved many of Fermat's other unproven margin-note claims, this claim, now known as Fermat's Last Theorem (FLT), was not proven until the 1990's (Wiles and Taylor.)

3. Pythagorean triples and FLT: cyclotomic integers

Just as the concept of irrational numbers was necessary to describe the lengths arising by elementary geometric constructions, the concept of *complex* numbers is needed to answer (other questions) and is helpful for generating Pythagorean triples. Recall that the imaginary unit i is defined to be $\sqrt{-1}$ and that a complex number is of the form $x + iy$, where x and y are real numbers.

Consider complex numbers of the form $a + bi$, where a and b are ordinary integers. Such complex numbers are called *Gaussian integers*, since it was Gauss who proved that they behave very much like the ordinary integers, particularly with respect to prime factorization. In the Gaussian integers we may factor $x^2 + y^2$:

$$x^2 + y^2 \;=\; (x - iy)(x + iy)$$

When x and y are relatively prime as ordinary integers, then (it turns out that) $x + iy$ and $x - iy$ are relatively prime as Gaussian integers. So if

$$(x - iy)(x + iy) \;=\; x^2 + y^2 \;=\; z^2$$

then $(x - iy)$ and $(x + iy)$ must be *squares* in the Gaussian integers (because they have no common prime factors!) In particular, there are ordinary integers u and v such that

$$x - iy \;=\; (u - iv)^2 \;=\; u^2 - 2iuv - v^2 \;=\; (u^2 - v^2) - 2iuv$$

Equating real and imaginary parts yields

$$x = u^2 - v^2 \quad \text{and} \quad y = 2uv$$

So we have recovered Euclid's formula for (primitive) Pythagorean triples again! (And this method was a lot quicker!)

In fact, using the *Eisenstein integers*

$$a + \left(\frac{-1+\sqrt{-3}}{2}\right)b \qquad \text{(where } a \text{ and } b \text{ are ordinary integers)}$$

(which also have unique prime factorization) one can prove FLT in the case $n = 3$. This success led others (e.g. Lamé, Kummer) to hope that understanding similar rings of "cyclotomic integers" could be used to prove FLT. However some of these rings do *not* have unique factorization. In order to repair this problem, Kummer began developing a theory of "ideal numbers," which enabled him to prove many cases of FLT, but not all.

4. The proof of FLT

The mathematics involved in the proof of FLT is very deep, and I will not attempt to explain it here. Instead I will just give a brief account of the major steps by which it was proven.

An *elliptic curve* is a curve of the form $y^2 =$ (cubic in x). The reason it is called *elliptic* is a bit complicated: it can be parametrized by elliptic functions, which are analogous to the inverse trigonometric functions for the circle.

In the 1950's and '60's Taniyama, Shimura, and Weil conjectured (TSW/"modularity conjecture") that every rational elliptic curve is *modular*, i.e. that it can be constructed from *modular functions.*

In the 1970's Hellegouarch studied the cubic curve $y^2 = (x - a^n)(x - b^n)(x - c^n)$ where $a^n + b^n = c^n$. A decade later Frey studied the same curve and showed that had some unusual properties, contrary to previous expectations.

Serre made some precise conjectures about these curves (almost) showing that TSW would imply FLT. In 1986, Ribet proved a the "epsilon conjecture" of Serre, thus reducing FLT to the (semi-stable case of) TSW.

In 1993, Wiles announced a proof of (the semi-stable case of) TSW, but it was found to have a flaw, which he subsequently repaired with the help of Taylor. Since that time the full modularity conjecture has been proven (Wiles-Taylor, Breuil-Conrad-Diamond.)

5. A philosophical note

At this point, perhaps you find yourself in a state of perplexity and you feel (as Meno did) like a torpedo has blown through your understanding of what mathematics is. You may be wondering whether these "algebraic integers" and "elliptic curves" aren't just the irrelevant imaginings of crazy mathematicians. If this is so, then remember that many others have thought that notions which we now consider obvious (like the existence of irrational numbers) have reacted in just the same way.

And perhaps, as Socrates says to Meno, being in a state of perplexity is better than being in a state of blissful ignorance, as long as one continues to strive for understanding: *We shall be better and braver and less helpless if we think that we ought to enquire, than we should have been if we indulged in the idle fancy that there was no knowing and no use in seeking to know what we do not know.*

6. Further reading

Perhaps you might be interested in some of the references I consulted when writing this talk.

I recommend John Stillwell's *Numbers and Geometry*, which situates "standard math" (arithmetic, geometry, trigonometry, etc.) in its historical context and discusses the significance of the insights which contributed to the development of these topics as well as the topics motivated by them. His *Elements of Number Theory* is also good, and I'd recommend it to anyone interested in number theory or any student of abstract algebra, since it explains the development of ring theory.

The Nova special called *The Proof* is a nice documentary (not too technical) about how Wiles proved FLT; it includes interviews with Wiles, Ribet, Shimura, and others. It provides a window into the world of research mathematics, and I recommend it to any of you interested in pursuing that. For a more mathematical discussion, you might look at Hellegouarch's *Invitation to the Mathematics of Fermat-Wiles.*

And for the philosophically minded, see http://classics.mit.edu/Plato/meno.html for the full text of Plato's *Meno.*