Diophantine equations, from Fermat to Wiles

Cutting Edge Lectures

McGill University

Montreal

October 2005

What is a number?

Mathematics is the study of **structure**.

Numbers are used to describe structure.

A *fluid, malleable concept*: rational, real, complex, quaternionic, p-adic, ...,

Discrete: *Counting*: $1, 2, 3, 4, 5, 6, 7, 8, 9, \ldots$, (number theory)

Continuous: *Measuring lengths* (geometry).

The Pythagorean credo ($\simeq$ 500 BCE):

All is number. (All lengths can be described by ratios of whole numbers.)

The square root of 2

Pythagorean Theorem: $a^2 + b^2 = c^2$.

If $a = b = 1$, then $c^2 = 2$.

Hippasus: The square root of 2 is *not* a ratio of whole numbers.

This discovery shook the foundations of the Pythagorean worldview. (And led to a divorce of number theory and geometry.)

Hippasus' Discovery

The square root of 2 is not a ratio of whole numbers; I.e., $a^2 = 2b^2$ has no solutions in *whole numbers* a and b.

Proof. (In the style of Fermat, 1601-1665).

$a^2 = 2b^2$, hence a^2 is even.

Hence a is even, i.e., $a = 2c$.

$4c^2 = 2b^2$, hence $b^2 = 2c^2$.

So $(b, c) = (b, a/2)$ is *another* **smaller** solution.

So $(a/2, b/2)$ is another solution.

So $(b/2, a/4)$ is another solution.

⋮ ⋮

But this cannot go on indefinitely: integers cannot be arbitrarily small!

Fermat championed this method of proof, which he called *infinite descent*.

Diophantine equations

A Diophantine equation is an equation (like $a^2 = 2b^2$) in which one is *only interested* in the whole number solutions.

Number Theory: the *art of solving Diophantine equations*.

Some examples:

Pythagorean equation: $x^2 + y^2 = z^2$.

Pell's equation: $x^2 - Dy^2 = 1$.

Fermat's equation: $x^n + y^n = z^n$.

Elliptic curves. $f(x, y, z) = 0$, all terms of f have degree 3.

The Pythagorean equation

$$x^2 + y^2 = z^2$$

Motivation: Right angles triangles with integer side lengths

Babylonian tablets (1900-1600 BCE) contain lists of *Pythagorean triples*.

Plimpton 322 (Columbia University)

$$(x, y, z) = (4961, 6480, 8161)$$

Parametric solution:

$$x = u^2 - v^2, \quad y = 2uv, \quad z = u^2 + v^2.$$

Essentially *all* solutions are obtained in this way, for suitable values of u and v.

Pell's Equation, d'après Brahmagupta (628 AD)

$$x^2 - Dy^2 = 1.$$

Motivation: if $(x, y) =$ solution, then $\frac{x}{y} \simeq \sqrt{D}$.

Replication rule (Brahmagupta).

$$(x_1, y_1) * (x_2, y_2) = (x_1x_2 + Dy_1y_2, x_1y_2 + y_1x_2).$$

Example: $(3, 2)$ is a solution to $x^2 - 2y^2 = 1$.

$$\begin{aligned} (3,2) * (3,2) &= (17,12), && 1.4166666... \\ (3,2) * (17,12) &= (99,70), && 1.4142857... \\ (3,2) * (99,70) &= (577,408), && 1.4142156... \end{aligned}$$

Bhaskara (1150 AD)

Problem: find the *initial* solution (x_0, y_0).

Example: Smallest solution to $x^2 - 61y^2 = 1$:

$$(x, y) = (1766319049, 226153980)$$

Bhaskara: a method for quickly finding the *smallest* solution to Pell's equation (*Chakravala*, cyclic method)

Write

$$\sqrt{D} = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots}}$$

$$\frac{x_n}{y_n} = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_n}}}.$$

For $n >> 0$, (x_n, y_n) is a solution to Pell's equation.

Fermat (1601-1665)

Fermat's result: Given $x^2 - Dy^2 = 1$, there *always exists* a smallest solution (x_0, y_0) from which all other solutions can be obtained by repeated aplication of the *replication rule*.

He *rediscovered* Bhaskara's method for finding (x_0, y_0).

Fermat's Challenge to Wallis: Find the smallest solution to $x^2 - 313y^2 = 1$.

Answer: it is ...

$$\begin{aligned} x_0 &= 32188120829134849, \\ y_0 &= 1819380158564160. \end{aligned}$$

Fermat's Challenge to Posterity

The equation

$$x^n + y^n = z^n$$

has no whole number solutions when $n \geq 3$.

This challenge was laid to rest by Andrew Wiles in 1995.

Wiles' proof occupies 130 pages and relies on earlier work of

Deligne ($\simeq$ 200 pages) **Langlands** ($\simeq$ 60 pages) **Mazur** ($\simeq$ 750 pages) **Ribet** ($\simeq$ 50 pages) **Serre** ($\simeq$ 400 pages) **Shimura** ($\simeq$ 200 pages) **Weil** ($\simeq$ 100 pages)

The proof would be hard to present from scratch in less than a thousand densely written pages, incorporating many key 20th Century ideas.

Elliptic curves

An elliptic curve is an equation $f(x, y, z) = 0$ where all the terms of f are of degree 3.

Example: $x^3 + y^3 = z^3$

Standard reductions:

Rational solutions to $f(x, y, 1) = 0$.

Elementary changes of variables:

$y^2 = x^3 + ax + b$, $\quad a$, b rational parameters.

Key fact: Like Pell's equation, elliptic curve equations possess a *replication rule*.

The replication rule

The replication rule for an elliptic curve

Motivation: The congruent number problem

Definition. An integer n is a *congruent number* if it is the area of a right-angled triangle with rational side lengths.

Elementary manipulations: n is a congruent number if and only if the elliptic curve

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

has a rational solution.

Problem. Given n, is it congruent?

An example

6 is a congruent number...

... and so is 157!

The recipe of Bhaskara?

A general recipe for efficiently solving elliptic curve equations is *not known.*

This is one of the seven "millenium prize problems" proposed by the Clay Institute in Cambridge Mass.

A tantalising approach

Find rational solutions by constructing appropriate *real* solutions.

Complex solutions

It even helps to consider complex solutions!

Complex numbers: $a + bi$, where $i^2 = -1$.

(Arise in electricity and magnetism, ...)

There is a natural "exponential" function

$$\exp : \{ \text{ complex numbers} \} \longrightarrow \{\text{complex solutions of } E\}.$$

Complex solutions to E are *easily parametrised*.
(Like Pythagorean triples!)

This does not seem very useful a priori...

A miraculous recipe

For each prime p let

$N_p := \#\{1 \leq x, y \leq p, \text{ where } p \text{ divides } f(x,y)\}$.

$a_p = p - N_p$.

Define a_n by

$$\prod_p (1 - a_p p^{-s} + p^{1-2s})^{-1} = \sum_n a_n n^{-s}.$$

Package these coefficients in a generating series:

$$H(z) = \sum_{n=1}^{\infty} \frac{a_n}{n} e^{2\pi i n z}.$$

A miraculous recipe, (cont'd)

Consider $H(a + b\sqrt{-d})$, where $a, b, d > 0$ are *rational*.

Fact: The *complex solutions* $\exp(H(a+b\sqrt{-d}))$ give rise to *loads of* rational solutions to E.

This is remarkable: A priori, these solutions belong to the "continuous" realm of geometry, or analysis, not the "discrete" realm of number theory.

Why does this work?

1. Modularity: The generating series $H(z)$ is related to a *modular form*, satisfying all kinds of "magical" properties. This is the content of Wiles' breakthrough!

2. The theory of "complex multiplication". Rationality properties of $f(a + b\sqrt{-d})$ when f is a modular form. (19th Century).

Dirichlet (1805-1859). "... un rapprochement magnifique entre deux branches de la science des nombres."

More miraculous recipes?

Since $\simeq$ 2000, it was observed **empirically** that the "miraculous recipe" is but one instance in a *broader scheme* for finding rational solutions to elliptic curve equations.

Problem. We still need to understand why these more general "miraculous recipes" work!

Why study Diophantine equations?

The excuse: Diophantine equations lead to **structures** that are *rich*, *complex*, and *intricate*.

Significant applications will *inevitably ensue*.

Areas of application of *elliptic curves* : cryptography, error-correcting codes, data compression, spam reduction protocols...

A dissenter:

> "The 'real' mathematics of the 'real' mathematicians [...] is almost wholly 'useless'... It is not possible to justify the life of any genuine professional mathematician on the ground of the 'utility' of his work."

G.H. Hardy, *A Mathematician's Apology*, 1940

The real answer

Diophantine equations lead to beautiful structures and patterns.

> "The mathematician's patterns, like the painter's or the poet's must be beautiful; the ideas, like the colours or the words must fit together in a harmonious way. Beauty is the first test: there is no permanent place in this world for ugly mathematics."

G.H. Hardy, *A Mathematician's Apology*