

Invitation to the Proof of Fermat's Last "Theorem"

William F. Hammond

<http://www.albany.edu/~hammond/>

University at Albany Math Club

March, 2011

Abstract

Invitation to the Proof of Fermat's Last "Theorem"

Fermat's Last "Theorem" (ca. 1637) was finally proved in the mid-1990s by using the study of plane cubic curves of the form

$$y^2 = (x - A)(x - B)(x - C)$$

where A, B, and C are distinct integers.

This talk will provide an overview of the main ingredients.

1 The Statement

There is no solution in positive integers x, y, z of the equation

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

for $n \geq 3$.

Note: There are infinitely many essentially different solutions when $n = 1, 2$.

2 Old History

- The statement is equivalent to the statement that there are no non-zero integers x, y, z satisfying $x^n + y^n = z^n$ for $n \geq 3$.
- The statement is equivalent to the statement that there are no rational points off the coordinate axes on the plane curve $x^n + y^n = 1$ for $n \geq 3$.
- For odd exponents $n \geq 3$ the statement is equivalent to the statement that there are no non-zero integers such that $x^n + y^n + z^n = 0$.

3 Old History (continued)

- If the theorem is true when the exponent is a given n , then it is certainly true when the exponent is a multiple of that value of n .
- The case where n is 3 or 4 can be handled within the realm of “elementary” number theory. (See, for example, the classic text of Hardy & Wright.)
- Any integer $n \geq 3$ not divisible by 4 must be divisible by an odd prime.
- It remains to prove the theorem when the exponent n is a prime $p \geq 5$.

4 A Solution leads to a Cubic Curve

Let $p \geq 5$ be prime.

Suppose there are non-zero integers a, b, c such that

$$a^p + b^p = c^p \quad .$$

Then the plane cubic curve

$$y^2 = x(x - a^p)(x + b^p) \quad .$$

is an *elliptic curve* “defined over” \mathbf{Q} — the Frey-Hellegouarch curve.

5 Cubic Curves

Over any field K , e.g., \mathbf{Q} , \mathbf{C} , or $\mathbb{F}_p = \mathbf{Z}/p\mathbf{Z}$, after a (projective) change of coordinates in K a non-singular cubic curve with at least one K -valued point may be brought into “generalized Weierstrass form”

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 ,$$

and over an algebraically closed field of characteristic $\neq 2, 3$ into an equation of the form

$$y^2 = (x - \lambda_1)(x - \lambda_2)(x - \lambda_3) .$$

The latter is a non-singular cubic when

$$\Delta = \left(\prod_{i < j} (\lambda_i - \lambda_j) \right)^2 \neq 0 .$$

Example: For the Frey-Hellegouarch curve

$$\Delta = (abc)^{2p} .$$

6 Cubics over the Complex Numbers

Given a non-singular cubic curve C ,

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 ,$$

with coefficients in \mathbf{C} and $\Delta \neq 0$, the set of all solutions (x, y) in \mathbf{C}^2 together with the “distinguished point at infinity” forms a compact Riemann surface of genus one — a torus.

7 The Projective Plane

For a given field K

$$\mathbf{P}^2(K) = K^2 \cup (\text{line at infinity})$$

where

$$\begin{aligned} \text{line at infinity} &= \{\text{classes of parallel lines}\} \text{ in } K^2 \\ &= \{\text{lines through } (0,0)\} \text{ in } K^2 \\ &= \{\text{slopes of lines}\} \cup (\infty) \\ &= K \cup (\infty) \end{aligned}$$

Each line contains one and only one point (its parallel class) on the line at infinity. The “distinguished point at infinity” is the parallel class of vertical lines.

8 A Line Meets a Cubic in 3 Points

Given a non-singular cubic curve C ,

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 ,$$

with coefficients in K , every line in K^2 passing through 2 points of C meets C in a third point, allowing for multiplicities.

Proof. Parameterize the line and get a cubic equation in the parameter with two known roots in K .

9 The Distinguished Point at Infinity

Given a non-singular cubic curve C ,

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 ,$$

with coefficients in K , the distinguished point at infinity in $\mathbf{P}^2(K)$ lies on C .

Proof. Introduce homogeneous coordinates $(x, y, z) \neq (0, 0, 0)$ in \mathbf{P}^2 where:

- $(x_1, y_1, z_1) \equiv (x_2, y_2, z_2)$ if and only if $(x_2, y_2, z_2) = t(x_1, y_1, z_1)$ for some scalar $t \neq 0$.
- $(x, y, 1)$ is a homogeneous triple for the affine point (x, y) .
- (x, y, z) is a homogenous triple for an affine point when $z \neq 0$.
- (x, y, z) represents a point on the line at infinity if $z = 0$.
- $(1, m, 0)$ represents “slope” m on the line at infinity.
- $(0, 1, 0)$ represents the “distinguished point at infinity”.

In homogeneous coordinates the curve C has the equation

$$y^2z + a_1xyz + a_3yz^2 = x^3 + a_2x^2z + a_4xz^2 + a_6z^3 .$$

In homogeneous coordinates the line at infinity has the equation $z = 0$.

The intersection of the line at infinity with C has the equation $x^3 = 0$. Thus, C meets the line at infinity “triple” in the distinguished point at infinity.

10 The Group Law

Given a non-singular cubic curve C ,

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 ,$$

with coefficients in K , there is a unique “algebraic” group law on the points of C in $\mathbf{P}^2(K)$ characterized by the two conditions

1. The group origin 0 is the distinguished point at infinity.
2. For three points P, Q, R of C one has $P + Q + R = 0$ if and only if P, Q, R lie on a line.

Note: Although the commutative law is obviously automatic here, it is not easy to check the associative law.

11 The Group Negative

For a given point (c, d) on the cubic curve

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 ,$$

its negative in the group law is the point (c, d') where d, d' are the two roots of

$$y^2 + (a_1c + a_3)y = c^3 + a_2c^2 + a_4c + a_6 ,$$

as a quadratic equation in y .

12 Elliptic Curves

- The non-singular cubic curves defined over K with at least one K -valued point are the “group objects” in the category of algebraic curves defined over K .
- For a curve in generalized Weierstrass form, the required K -valued point may always be taken to be the distinguished point at infinity.
- These are called elliptic curves.
- When $K = \mathbf{Q}$, much is known about them.
- *Modular forms* — objects associated with hyperbolic geometry — provide a dictionary for elliptic curves defined over \mathbf{Q} .
- The Frey-Hellegouarch curve cannot be in that dictionary.

13 The mod ℓ reduction of an elliptic curve

Let E be an elliptic curve of the form

$$y^2 = (x - A)(x - B)(x - C)$$

where A, B, C are distinct integers. When ℓ is a prime not dividing Δ (the square of the product of the root differences), E determines also a curve E_ℓ defined over the finite field $\mathbb{F}_\ell = \mathbf{Z}/\ell\mathbf{Z}$. E_ℓ is non-singular when ℓ is not a factor of Δ .

For our purposes, i.e., in the case of the Frey-Hellegouarch curve, the *conductor* N of E may be defined to be

$$N = \prod_{\ell|\Delta} \ell,$$

the square-free part of Δ .

Let c_ℓ be defined by

$$c_\ell = 1 - |E(\mathbb{F}_\ell)| + \ell$$

when $\ell \nmid N$. Here $|E(\mathbb{F}_\ell)|$ denotes the number of points of E_ℓ in the field \mathbb{F}_ℓ .

c_ℓ is defined in a slightly more complicated way for each of the finitely many primes ℓ dividing N .

14 The L-series of E

One defines the “L-series” of E by forming the Euler product, indexed by primes ℓ as follows:

$$L(E, s) = \prod_{\ell \nmid N} \frac{1}{1 - c_\ell \ell^{-s}} \prod_{\ell \mid N} \frac{1}{1 - c_\ell \ell^{-s} + \ell^{1-2s}}$$

Expanding the product, one obtains a Dirichlet series

$$L(E, s) = \sum_{k=1}^{\infty} \frac{c_k}{k^s},$$

which converges for $\text{Re}(s) > 3/2$

Series of this type have been seen in other contexts.

15 Isometries of the Upper-Half Plane

Let \mathbf{H} be

$$\mathbf{H} = \{\tau \in \mathbf{C} \mid \text{Im}(\tau) > 0\} \quad .$$

The group $G = \text{SL}_2(\mathbf{R})$ operates via

$$M \cdot \tau = \frac{a\tau + b}{c\tau + d}, \quad M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad a, b, c, d \in \mathbf{R}, \quad ad - bc = 1$$

$G/\{\pm 1\}$ is the group of isometries (distance-preserving analytic maps) of \mathbf{H} relative to the Poincaré metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2}, \quad \text{for } \tau = x + iy \in H \quad .$$

(This is the connection with “hyperbolic geometry”.)

16 Family of Elliptic Curves over \mathbf{C}

Let G_w denote the Eisenstein series

$$G_w(\tau) = \text{const} \cdot \sum_{(m,n) \in \mathbf{Z}^2 - \{(0,0)\}} \frac{1}{(m\tau + n)^w},$$

which converges normally for all $\tau \in \mathbf{H}$, $w \geq 4$.

$G_w(\tau)$ is not identically 0 for even $w \geq 4$, while it is self-cancelling for odd w .

For given τ with $g_4(\tau) = 60G_4(\tau)$, $g_6(\tau) = 140G_6(\tau)$ the equation

$$y^2 = 4x^3 - g_4(\tau)x - g_6(\tau)$$

gives rise to a cubic curve C_τ in classical Weierstrass form.

Every elliptic curve defined over \mathbf{C} occurs this way, and one has

$$C_{\tau'} \cong C_\tau \Leftrightarrow \tau' = M \cdot \tau \text{ for } M \in \text{SL}_2(\mathbf{Z}) \quad .$$

Thus, over \mathbf{C}

$$\{\text{isomorphism classes of elliptic curves}\} \cong \mathbf{H}/\text{SL}_2(\mathbf{Z})$$

17 Modular Forms

The Eisenstein series are examples of modular forms: complex-valued holomorphic functions f in \mathbf{H} satisfying

$$f(M \cdot \tau) = (c\tau + d)^w f(\tau)$$

for

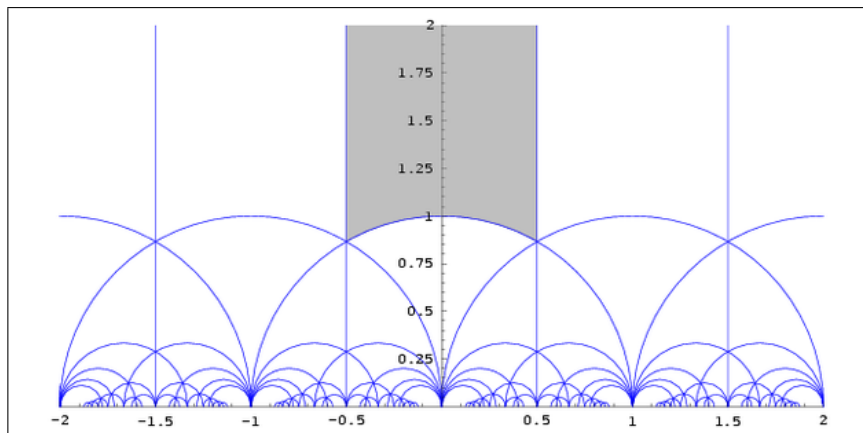
$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma, \quad \tau \in \mathbf{H} \quad .$$

where Γ is $\mathrm{SL}_2(\mathbf{Z})$ or a subgroup of finite index in $\mathrm{SL}_2(\mathbf{Z})$.

- The integer w is the *weight* of f .
- G_w is a modular form of weight k .
- A modular form is, more or less, a holomorphic section of a “line bundle” on the quotient space \mathbf{H}/Γ .
- Modular forms are also required to be “holomorphic at cusps”, i.e., approach a finite limit at a “cusp” (see below).

18 Action of $\mathrm{SL}_2(\mathbf{Z})$ on \mathbf{H}

The action of $\Gamma = \mathrm{SL}_2(\mathbf{Z})$ on \mathbf{H} is portrayed in this picture:



(Wikipedia image licensed under GFDL)

- The gray area is a fundamental domain. It has infinite extent.
- \mathbf{H}/Γ is non-compact.

19 Cusps Compactify the Quotient

Let Γ be a subgroup of finite index in $\Gamma_0(1) = \mathrm{SL}_2(\mathbf{Z})$.

- \mathbf{H}/Γ “covers” $\mathbf{H}/\Gamma_0(1)$
- Γ operates on $\mathbf{H}^* = \mathbf{H} \cup \mathbf{Q} \cup \{\infty\}$.
- For $\Gamma = \Gamma(1)$ the orbit of ∞ is $\mathbf{Q} \cup \{\infty\}$.
- For general Γ the number of orbits in $\mathbf{Q} \cup \{\infty\}$ is finite.
- \mathbf{H}^*/Γ compactifies \mathbf{H}/Γ by adjoining the finitely many “cusps”.

20 Cusp Forms

Let Γ be a subgroup of finite index in $\Gamma_0(1) = \mathrm{SL}_2(\mathbf{Z})$. A modular form for Γ is a *cuspidal form* if its limiting value at each cusp of Γ is 0.

Example: For $\Gamma = \Gamma_0(1) = \mathrm{SL}_2(\mathbf{Z})$ the modular form

$$\lambda(\tau) = g_4(\tau)^3 - 27g_6(\tau)^2$$

is a cusp form of weight 12 — the smallest weight of a cusp form for $\Gamma_0(1)$.

21 The Groups $\Gamma_0(N)$

Let $N \geq 1$ be a positive integer. The group $\Gamma_0(N)$ is given by

$$\left\{ M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \mid c \equiv 0 \pmod{N} \right\} .$$

In particular

$$M_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma_0(N) \text{ for all } N \geq 1 .$$

If f is a modular form, then

$$f(M_1 \cdot \tau) = f(\tau + 1) = f(\tau)$$

is periodic, so has a Fourier expansion

$$f(\tau) = \sum_{k \in \mathbf{Z}} c_k e^{2\pi i k \tau} .$$

Because f is holomorphic at the cusp ∞ one has $c_k = 0$ for $k < 0$, and if f is a cusp form $c_0 = 0$ so that then

$$f(\tau) = \sum_{k=1}^{\infty} c_k e^{2\pi i k \tau} .$$

N is called the *level*.

22 The Dirichlet Series

There are certain operators, called Hecke operators $\{T_w(m)\}_{m \geq 1}$, that act semi-simply on the space of cusp forms for $\Gamma_0(N)$ not coming from levels dividing N . The structure of the algebra of these operators shows that if

$$f(\tau) = \sum_{k=1}^{\infty} c_k e^{2\pi i k \tau}$$

is a cusp form of **weight 2** that is a simultaneous eigenform of these operators then the corresponding Dirichlet series

$$\varphi_f(s) = \sum_{k=1}^{\infty} \frac{c_k}{k^s}$$

has an Euler product expansion just like the Euler product that is the L-function of an elliptic curve defined over \mathbf{Q} :

$$\varphi_f(s) = \prod_{\ell|N} \frac{1}{1 - c_\ell \ell^{-s}} \prod_{\ell \nmid N} \frac{1}{1 - c_\ell \ell^{-s} + \ell^{1-2s}}$$

23 Cusp Forms of Weight 2 on $\Gamma_0(N)$

A cusp form f of weight 2 for $\Gamma_0(N)$ is essentially a regular differential on the quotient $X_0(N) = \mathbf{H}^*/\Gamma_0(N)$. When f , not coming from levels dividing N , is an eigenform of the Hecke operators, it determines in a straightforward way a 1-dimensional quotient variety of the Jacobian variety $J_0(N)$ of $X_0(N)$,

$$X_0(N) \longrightarrow J_0(N) \longrightarrow E_f$$

which quotient is an elliptic curve E_f defined over \mathbf{Q} with conductor N , and, therefore a regular map – the modular parametrization of E_f – from $X_0(N)$ to E_f with the property that the unique (up to a constant) regular differential on E_f pulls back to the differential on $X_0(N)$ determined by f .

24 The Dictionary for Elliptic Curves over \mathbf{Q}

- Since the mid 20th century one has known that a cusp form f of weight 2 for $\Gamma_0(N)$, not also residing at a level dividing N , that is a simultaneous eigenform for the Hecke operators gives rise to an elliptic curve E_f defined over \mathbf{Q} with conductor N .
- The L-function of E_f is the Dirichlet series $\varphi_f(s)$ associated with f .
- The “Modular Curve Conjecture”, which originated in the mid 20th century, is that every elliptic curve defined over \mathbf{Q} is isogenous to one of those obtained from such a cusp form. (Isogenous elliptic curves share L-functions.)
- In the mid 1980s it was shown using the theory of representations of the Galois group of the field of all algebraic numbers that the dictionary for elliptic curves defined over \mathbf{Q} provided by the modular curve conjecture (and the extensive knowledge of modular forms) could not possibly contain the Frey-Hellegouarch curve.
- In the 1990s the “Modular Curve Conjecture” was proved.
- Fermat’s Last Theorem is a corollary of that above.

25 Dictionary Trivia

- 11 is the smallest value of N for which there is a non-zero cusp form of weight 2 for the group $\Gamma_0(N)$. In this case the dimension of the space of cusp forms is 1. There are 3 non-isomorphic but isogenous elliptic curves with conductor 11:

$$\begin{aligned}y^2 + y &= x^3 - x^2 - 10x - 20 \\y^2 + y &= x^3 - x^2 - 7820x - 263580 \\y^2 + y &= x^3 - x^2\end{aligned}$$

- The Cremona database — an encoding of the dictionary — has been built into *Sage* (<http://www.sagemath.org/>). Documentation for its use may be found at <http://www.sagemath.org/doc/reference/sage/databases/cremona.html>.
- The smallest conductor having more than 1 isogeny class is 26, which has 2.
- The smallest conductor having more than 2 isogeny classes is 57, which has 3.
- There are 38402 isogeny classes with conductors smaller than 10000.

26 For More Information

G. Cornell, J. Silverman, & G. Stevens,
Modular Forms and Fermat's Last Theorem,
Springer, 1997

— the record of an instructional conference held at Boston University in August, 1995

27 Acknowledgement

The XHTML + MATHML version of these slides uses W3C's *Slidy* by Dave Raggett, a JavaScript/CSS package for sizing and flow control of an HTML or XHTML slide show.

(The slides were generated in a non-standard fashion from GELLMU source.)