The Modularity Theorem

Bogdan Grechuk • 2 Jan 2026

As announced in my previous blog post, I am beginning a series of posts devoted to my favorite theorems of the 21st century. In the category "Between the Centuries" in Number Theory, my personal favorite is the *Modularity Theorem*.

A fundamental role in modern number theory is played by curves E of the form

$$y^2 = x^3 + ax + b, \tag{1}$$

where $a, b \in \mathbb{Z}$ and the discriminant

$$\Delta_E := -16(4a^3 + 27b^2)$$

is nonzero. The equation (1) defines a *non-singular elliptic curve over* $\mathbb{Q}$ *in Weierstrass form*, or simply an *elliptic curve*.

Elliptic curves have proved to be extraordinarily powerful tools: they have played a decisive role in the solution of many problems whose original formulations bore no apparent connection to them. The most celebrated example is the proof of *Fermat's Last Theorem*.

Around 1637, Pierre de Fermat asserted—famously without proof—that for any integer $n \geq 3$ there are no positive integers x, y, z satisfying

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

This claim, later known as Fermat's Last Theorem, remained the most notorious open problem in mathematics for more than three centuries.

In 1955, Taniyama and Shimura initiated a remarkable connection between elliptic curves and certain analytic functions defined on the complex upper half-plane

$$\mathbb{H} := \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}.$$

These functions are now known as *modular forms*.

A modular form of weight $k \geq 0$ and level N is a holomorphic function $f : \mathbb{H} \to \mathbb{C}$ satisfying:

- the transformation law

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z), \tag{2}$$

for all integers a, b, c, d with $ad - bc = 1$ and $N \mid c$;

- a growth condition, stating that for every $a, b, c, d \in \mathbb{Z}$ such that $ad - bc = 1$,

$$(cz+d)^{-k} f\left(\frac{az+b}{cz+d}\right)$$

remains bounded as $\mathrm{Im}(z) \to \infty$.

These conditions imply that f admits a Fourier expansion

$$f(z) = \sum_{n=0}^{\infty} c_n \, e^{2\pi i n z}, \qquad z \in \mathbb{H}, \tag{3}$$

with complex coefficients $c_n = c_n(f)$.

A central theme in the arithmetic of elliptic curves is the study of the number $N_p = N_p(E)$ of solutions to (1) modulo a prime p, that is, the number of $\mathbb{F}_p$-points on E.

In 1936, Hasse (Hasse 1936), confirming a conjecture of Artin from 1924, proved the celebrated estimate

$$p - 2\sqrt{p} \le N_p(E) \le p + 2\sqrt{p}. \tag{4}$$

Equivalently, one defines the quantity

$$a_p = a_p(E) := p - N_p(E), \tag{5}$$

known as the *trace of Frobenius* at p, which satisfies $|a_p| \le 2\sqrt{p}$.[1]

An elliptic curve E over $\mathbb{Q}$ is called *modular* if there exists a modular form $f = f(E)$ such that, for every prime p not dividing Δ_E, the equality

$$a_p(E) = c_p(f)$$

holds between the trace of Frobenius of E and the p-th Fourier coefficient of f.

Taniyama and Shimura conjectured that *every elliptic curve over $\mathbb{Q}$ is modular*. This statement, now known as the *Taniyama–Shimura conjecture* or the *modularity conjecture*, was strikingly unexpected: the sequences $\{a_p(E)\}$ and $\{c_p(f)\}$ arise from entirely different worlds—arithmetic geometry on one side and complex analysis on the other—and their agreement appears almost miraculous.

In 1986, Ribet (Ribet 1990), proving a conjecture of Serre, showed that if Fermat's Last Theorem were false, then there would exist an elliptic curve—now called the *Frey curve*—which is not modular. In 1995, Wiles (Wiles 1995) proved the modularity conjecture for a large class of elliptic curves, including the Frey curve. Together, these results established Fermat's Last Theorem.

While Wiles's work stands as one of the crowning achievements of 20th-century mathematics, it did not settle the modularity conjecture in full generality. This final step was achieved in 2001 by Breuil, Conrad, Diamond, and Taylor (Breuil et al. 2001).

Theorem 1 (Modularity Theorem) *Every elliptic curve E over $\mathbb{Q}$ is modular. More precisely, E corresponds to a modular form of weight $k = 2$ and level Δ_E.*

The authors of (Breuil et al. 2001) proved a substantially stronger version of Theorem 1, establishing additional structural properties of the associated modular form. For the purposes of most applications, however, the formulation above—communicated to the author by Dan Fretwell—captures the essential content while remaining easy to state.

The proof of Theorem 1 builds on Wiles's groundbreaking ideas, extending and refining them to treat the remaining cases. Since 2001, the Modularity Theorem and the methods developed for its proof have become central tools in number theory, with far-reaching consequences well beyond their original scope.

References

Breuil, Christophe, Brian Conrad, Fred Diamond, and Richard Taylor. 2001. "On the Modularity of Elliptic Curves over Q: Wild 3-Adic Exercises." *J. Amer. Math. Soc.* 14 (4): 843–939. https://doi.org/10.1090/S0894-0347-01-00370-8.

Hasse, Helmut. 1936. "Zur Theorie Der Abstrakten Elliptischen Funktionenkörper III. Die Struktur Des Meromorphismenrings. Die Riemannsche Vermutung." *J. Reine Angew. Math.* 1936 (175): 193–208.

Ribet, Kenneth A. 1990. "On Modular Representations Arising from Modular Forms." *Invent. Math.* 100 (1): 431–76.

Wiles, Andrew. 1995. "Modular Elliptic Curves and Fermat's Last Theorem." *Ann. Of Math.* 141 (3): 443–551.

---

1. Often one includes the point at infinity among the solutions modulo p. In that convention, $a_p = p + 1 - M_p$, where M_p is the total number of points.↩