MODULAR ELLIPTIC CURVES AND DIOPHANTINE PROBLEMS

by Dorian Goldfeld[1]

§1. Introduction:

Let E be an elliptic curve, defined over $\mathbb{Q}$, given in Weierstrass normal form

$$\begin{aligned} E : y^2 &= x^3 - ax - b \\ &= (x - e_1)(x - e_2)(x - e_3). \end{aligned}$$

The discriminant of E is defined to be $D = (e_1 - e_2)^2(e_1 - e_3)^2(e_2 - e_3)^2$. Two elliptic curves given in Weierstrass normal form will be isomorphic if and only if they are equivalent under a rational transformation of type $x \mapsto u^2x,\ y \mapsto u^3y$ with $u \in \mathbb{Q}$, and u unequal to 0. Under this transformation a is transformed to $u^{-4}a$ and b is transformed to $u^{-6}b$. Similarly, D is transformed to $u^{-12}D$.

We say E is in minimal Weierstrass normal form or is a minimal Weierstrass model over $\mathbb{Q}$ if among all isomorphic Weierstrass models for E (with $a, b \in \mathbb{Z}$) we have that D is minimized.

If the cubic $x^3 - ax - b = (x - e_1)(x - e_2)(x - e_3)$ has three distinct real roots, then the real points of E (denoted $E(\mathbb{R})$) has two nonsingular connected components which are symmetric with respect to the x-axis. Although $E(\mathbb{R})$ is nonsingular, it may very well happen that $E(\mathbb{F}_p)$ (where $\mathbb{F}_p$ is the finite field of p elements) is singular. It is not hard to see that this can only happen for primes $p|D$, and such primes are called primes of bad reduction. A measure for the amount of bad reduction is given by the conductor of the elliptic curve. The conductor is denoted by the symbol N and is defined as follows:

$$N = \prod_{p|D} p^{e(p)}$$

where for p unequal to 2 or 3, $e(p) = 1$ if the singularity is a node, curve with two distinct tangent lines at the singular point, while $e(p) = 2$ if the singularity is a cusp, curve with one tangent at the singular point, and in the remaining cases of $p = 2, 3$, $e(p)$ is absolutely bounded. An elliptic curve is said to be semistable if it never has bad reduction of cuspidal type, and in this case N is always the squarefree part of D.

In a remarkable series of papers [F1], [F2], G. Frey constructed minimal semistable elliptic curves over $\mathbb{Q}$. Let me briefly describe Frey's construction. Let $A, B, C \in \mathbb{Z}$ with $A \equiv 0(32)$, $B \equiv 1(4)$, $(A, B) = 1$, and $A + B + C = 0$. Consider the elliptic curve

$$E_{A,B} : \ y^2 = x(x - A)(x + B).$$

A normal Weierstrass form for E is given by

$$\tilde{E}_{A,B} : \ y^2 = x^3 - \alpha x + \beta \qquad (1)$$

[1] This work was done while the author was partially supported by a grant from the Vaughn Foundation.

where we have

$$\alpha = \frac{1}{3}\Big(A^2 + B^2 + AB\Big), \quad \beta = \frac{1}{27}\Big(A + B\Big)\Big(2A^2 + 2B^2 + 5AB\Big),$$

and $\alpha, \beta \in \mathbb{Z}$ if and only if $A \equiv B(3)$. Frey shows that this curve is semistable. Moreover, in the case $A \equiv B(3)$, since $(\alpha, \beta) = 1$, $\tilde{E}_{A,B}$ is in minimal Weierstrass form with discriminant $A^2B^2C^2$. On the other hand, if $A \not\equiv B(3)$, then the simple transformation $x \mapsto \frac{1}{9}x, \ \ y \mapsto \frac{1}{27}y$, gives a minimal Weierstrass normal form with discriminant $3^{12}A^2B^2C^2$. Note that our definition of minimal Weierstrass normal form is different from the usual notion of minimal model over $\mathbb{Z}$. Frey shows that a minimal model for $E_{A,B}$ over $\mathbb{Z}$ is given by the curve

$$y^2 + xy = x^3 + \frac{A - B - 1}{4}x^2 - \frac{AB}{16}x$$

with minimal discriminant $A^2B^2C^2/256$.

A surprisingly novel idea of Frey is to suggest that if the Fermat equation

$$u^p + v^p + w^p = 0$$

has a nontrivial solution in rational integers u, v, w for $p > 2$ then the elliptic curve (1) with $A = u^p, \ B = v^p, \ C = w^p$ cannot exist as a minimal Weierstrass model. Using this approach and earlier work of Mazur [M2], and Serre [S1], [S2], Ribet [R] has recently shown that Fermat's last theorem would follow from the conjecture of Taniyama and Weil which is described in the next section. I shall not discuss Ribet's theorem in this article, but focus instead on another approach of Frey [F2] based on a conjecture of Szpiro [Szp1], [Szp2], (1983).

Let

$$E: \ \ y^2 = x^3 - ax - b$$

be an elliptic curve with $a, b \in \mathbb{Z}$, D nonzero, in minimal Weierstrass form. Let N be the conductor of E.

Conjecture(1) (Szpiro): *There exists an absolute constant κ (independent of N, D) such that*

$$D \ \le \ N^\kappa.$$

A stronger form of this conjecture states that if E is also semistable then

Conjecture(2) (Szpiro): *For every $\epsilon > 0$ there exists a constant $c(\epsilon)$ depending only on ϵ such that*

$$D \ \le \ c(\epsilon)N^{6+\epsilon}.$$

Applying this to the Frey curve (1), for example, yields the inequality

$$|ABC|^2 \ \le \ c(\epsilon) \prod_{p|ABC} p^{6+\epsilon},$$

and this proves Fermat's last theorem for all sufficiently large exponents p. On the basis of the above example, Masser and Osterlé [Ost] (1985) conjectured the following.

Conjecture(3); *For rational integers A, B, C with $A+B+C=0$*

$$\sup(|A|,|B|,|C|) \ll \prod_{p|ABC} p^{1+\epsilon},$$

where the $\ll$-constant depends at most on $\epsilon > 0$.

In fact, conjecture(3) with $\sup(|A|,|B|,|C|)$ replaced by $|ABC|^{\frac{1}{3}}$ follows from conjecture (2). We also remark that conjecture (1) should hold over any number field with a constant κ depending at most on the field. Recently, Hindry and Silverman [H-S] showed that Lang's conjecture on the lower bound for the height of non-torsion points on an elliptic curve over a number field follows from conjecture (1), and more recently, Frey [F3], under the assumption of conjecture (1) gave a bound for the order of a torsion point on an elliptic curve defined over a number field. If Szpiro's conjecture is proven, this would generalize an unconditional result of Mazur [M1] which says that a torsion point on an elliptic curve defined over $\mathbb{Q}$ can be of order at most twelve.

§2. The conjecture of Taniyama and Weil

We now consider the elliptic curve

$$E: \; y^2 = 4x^3 - ax - b \tag{2}$$

where for simplicity we assume that $4x^3 - ax - b = 4(x-e_1)(x-e_2)(x-e_3)$ and the three roots $e_1 < e_2 < e_3$ are real.The periods of E (denoted Ω_1, Ω_2) are defined by the integrals

$$\Omega_1 = 2\int_{e_3}^{+\infty} \frac{dx}{\sqrt{4x^3 - ax - b}}$$
$$\Omega_2 = 2\int_{e_2}^{e_3} \frac{dx}{\sqrt{4x^3 - ax - b}}$$

where Ω_1 is real and Ω_2 is pure imaginary. Let $D = a^3 - 27b^2$ be the discriminant of E. It is well known that E can be parametrized by doubly periodic functions

$$x = \wp(z)$$
$$y = \wp'(z)$$

where

$$\wp'(z) = -2\sum_{m,n\in\mathbb{Z}} \frac{1}{(z + m\Omega_1 + n\Omega_2)^3},$$

and this is just the generalization of the well known parametrization of the circle $x^2 + y^2 = 1$ by the trigonometric functions $x = \cos z, \;\; y = \sin z$.

The Taniyama-Weil conjecture in its simplest form states that every elliptic curve E defined over $\mathbb{Q}$, in minimal form and with conductor N, can be parametrized by modular functions for the group (see [M-Sw])

$$\Gamma_o(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; a,b,c,d \in \mathbb{Z},\; ad - bc = 1,\; c \equiv 0 \pmod{N} \right\}.$$

That is to say there exist meromorphic functions $\alpha(z), \beta(z)$ with z in the upper half plane satisfying

$$\alpha\left(\frac{az+b}{cz+d}\right) = \alpha(z)$$

$$\beta\left(\frac{az+b}{cz+d}\right) = \beta(z).$$

for all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_o(N)$. Moreover, the curve

$$y^2 = 4x^3 - ax - b$$

can be parametrized by

$$x = \alpha(z)$$

$$y = \beta(z).$$

We shall now explicitly construct $\alpha(z)$, $\beta(z)$, assuming they exist.

Let

$$f(z) = \sum_1^\infty a(n) e^{2\pi i n z}$$

be a cusp form of weight 2 for $\Gamma_o(N)$ so that

$$f(\frac{az+b}{cz+d}) = (cz+d)^2 f(z).$$

We assume that f is normalized so that $a(1) = 1$, $a(n) \in \mathbb{Z}$ for $n \geq 1$, and that

$$a(mn) = a(m)a(n)$$

for $(m, n) = 1$.

Let $X_o(N)$ be the modular curve of the compactified Riemann surface obtained from factoring the upper half plane by $\Gamma_o(N)$. By a theorem of Shimura [Sh], there exists an elliptic curve E which we may take to be (2) and a covering map ϕ, normalized so that $\phi(i\infty) = 0$,

$$\begin{array}{c} X_o(N) \\ \Big\downarrow \phi \\ E \end{array}$$

so that $f(z)dz$ is the pullback under ϕ of a differential one-form on E.

Let

$$F(\tau) = -2\pi i \int_\tau^{i\infty} f(z)\, dz$$

$$= \sum_1^\infty \frac{a(n)}{n} e^{2\pi i n \tau}$$

be the antiderivative of f. For $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_o(N)$ let us consider the Shimura map

(3) $$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \mapsto \quad F\left(\frac{a\tau+b}{c\tau+d}\right) - F(\tau).$$

By the fundamental theorem of calculus

$$\frac{\partial}{\partial\tau}\{F\left(\frac{a\tau+b}{c\tau+d}\right)-F(\tau)\} \quad = \quad 0,$$

so the right side of (3) is independent of τ. We now define

$$H\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) \quad = \quad F\left(\frac{a\tau+b}{c\tau+d}\right)-F(\tau)$$

to be the Shimura map.

Since for $\alpha_1, \alpha_2 \in \Gamma_o(N)$ we have

$$\begin{aligned} H(\alpha_1\alpha_2) &= F(\alpha_1(\alpha_2\tau)) - F(\alpha_2\tau) + F(\alpha_2\tau) - F(\tau) \\ &= H(\alpha_1) + H(\alpha_2) \end{aligned}$$

we see that H is a homomorphism of $\Gamma_o(N)$. In fact if the pullback $\phi^*(f(z)dz)$ is the standard differential on E then

$$H(\alpha) = 2\pi i \int_{\tau}^{\alpha\tau} f(z)\,dz$$

must lie in the homology of $X_o(N)$ and hence in the homology of E. It follows that H is a homomorphism from $\Gamma_o(N)$ onto the lattice

$$\Lambda = \{m\Omega_1 + n\Omega_2 \mid m, n \in \mathbb{Z}\}$$

of periods of E which is just an abelian group of rank 2 isomorphic to $\mathbb{Z}\times\mathbb{Z}$.

We can now give the desired parametrization of $E : y^2 = 4x^3 - ax - b$. Let us define

$$\alpha(z) = \wp(F(z)) = \wp\left(\sum_{n=1}^{\infty}\frac{a(n)}{n}e^{2\pi inz}\right)$$

$$\beta(z) = \wp'(F(z)) = \wp'\left(\sum_{n=1}^{\infty}\frac{a(n)}{n}e^{2\pi inz}\right),$$

where $\wp$ is the Weierstrass $\wp$-function. We have

$$\begin{aligned} \alpha\left(\frac{az+b}{cz+d}\right) &= \wp\left(F\left(\frac{az+b}{cz+d}\right)\right) \\ &= \wp\left(F(z) + H\left(\begin{pmatrix} a\,b \\ c\,d \end{pmatrix}\right)\right) \\ &= \wp\big(F(z)\big) \\ &= \alpha(z) \end{aligned}$$

since $H\left(\left(\begin{smallmatrix} a\,b \\ c\,d \end{smallmatrix}\right)\right) \in \Lambda$. Similarly for $\beta(z)$.

§3. Properties of Shimura maps:

The Shimura map $H : \Gamma_o(N) \to \Lambda$ as defined in the previous section satisfies the following properties:

Property (1): *H is a homomorphism from $\Gamma_o(N)$ onto the period lattice Λ of the elliptic curve E.*

Property (2): *For $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_o(N)$, we have $H\left(\left(\begin{smallmatrix} a & -b \\ -c & d \end{smallmatrix}\right)\right) = \overline{H\left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right)}$.*

Proof: Let $\sigma = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$ with $i = \sqrt{-1}$. Then we have

$$\sigma \begin{pmatrix} a & b \\ c & d \end{pmatrix} \sigma^{-1} = \begin{pmatrix} ai & bi \\ -ci & -di \end{pmatrix} \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}.$$

Since the Fourier coefficients of f are real it follows that

$$F(\sigma\overline{z}) = F(\sigma^{-1}\overline{z}) = F(-\overline{z}) = \overline{F(z)}.$$

Hence, replacing τ by $\sigma\overline{\tau}$, we have

$$\begin{aligned} H\left(\begin{pmatrix} a & -b \\ -c & d \end{pmatrix}\right) &= F\left(\sigma \begin{pmatrix} a & b \\ c & d \end{pmatrix} \sigma^{-1}\tau\right) - F(\tau) \\ &= F\left(\sigma \begin{pmatrix} a & b \\ c & d \end{pmatrix} \overline{\tau}\right) - F\left(\sigma\overline{\tau}\right) \\ &= \overline{H\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right)}. \end{aligned}$$

Property (3): *For each positive squarefree integer N, there exists $\epsilon_N = \pm 1$ such that for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_o(N)$, we have*

$$H\left(\begin{pmatrix} d & -\frac{c}{N} \\ -bN & a \end{pmatrix}\right) = \epsilon_N H\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right).$$

Proof: Let $\omega = \begin{pmatrix} 0 & \frac{1}{\sqrt{N}} \\ -\sqrt{N} & 0 \end{pmatrix}$ so that

$$\omega \begin{pmatrix} a & b \\ c & d \end{pmatrix} \omega^{-1} = \begin{pmatrix} d & -\frac{c}{N} \\ -bN & a \end{pmatrix}.$$

It follows that

$$\begin{aligned} H\left(\begin{pmatrix} d & -\frac{c}{N} \\ bN & a \end{pmatrix}\right) &= H\left(\omega \begin{pmatrix} a & b \\ c & d \end{pmatrix} \omega^{-1}\right) \\ &= F\left(\omega \begin{pmatrix} a & b \\ c & d \end{pmatrix} \omega^{-1}\tau\right) - F(\tau) \\ &= L + M + N \end{aligned}$$

where

$$\begin{aligned}
L &= F\left(\omega\begin{pmatrix} a & b \\ c & d\end{pmatrix}\omega^{-1}\tau\right) - \epsilon_N F\left(\begin{pmatrix} a & b \\ c & d\end{pmatrix}\omega^{-1}\tau\right) \\
M &= \epsilon_N F\left(\begin{pmatrix} a & b \\ c & d\end{pmatrix}\omega^{-1}\tau\right) - \epsilon_N F(\omega^{-1}\tau) \\
N &= \epsilon_N F(\omega^{-1}\tau) - F(\tau).
\end{aligned}$$

By the functional equation $F(\tau) = \epsilon_N F(\omega\tau)$, we have $L = 0$, and $N = 0$. The result follows.

Property (4): *Let* $\sigma_p = \left(\begin{smallmatrix} p & 0 \\ 0 & 1\end{smallmatrix}\right)$ *and* $\sigma_j = \left(\begin{smallmatrix} 1 & j \\ 0 & p\end{smallmatrix}\right)$ *for* $j = 0, 1, ...(p-1)$. *Assume that* $\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)$, $\sigma_k \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\sigma_k^{-1} \in \Gamma_o(N)$ *for* $k = 0, 1, ..., p$. *(This will be the case if* $p|b$, $p|c$, *and* $p|(d-a)$.*) Then for* p *a rational prime not dividing* N *we have*

$$\sum_{k=0}^{p} H\left(\sigma_k \begin{pmatrix} a & b \\ c & d\end{pmatrix}\sigma_k^{-1}\right) = a(p) H\left(\begin{pmatrix} a & b \\ c & d\end{pmatrix}\right)$$

where $a(p) = p^{th}$ *Fourier coefficient of* $f(z)$.

Proof: We make use of the properties of the Hecke operator $T_p = \sum_{k=0}^{p} \sigma_k$ and the fact that the differential one form $f(z)dz$ is an eigenfunction of T_p with eigenvalue $a(p)$

$$T_p(f(z)dz) = a(p)f(z)dz.$$

From the definition of H we see that

$$\sum_{k=0}^{p} H\left(\sigma_k \begin{pmatrix} a & b \\ c & d\end{pmatrix}\sigma_k^{-1}\right) = \sum_{k=0}^{p}\left[\int_{\sigma_k\alpha\tau_o}^{i\infty} f(z)\,dz \;-\; \int_{\sigma_k\tau_o}^{i\infty} f(z)\,dz\right]$$

after putting $\alpha = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)$, and $\tau_o = {\sigma_k}^{-1}\tau$. It follows that

$$\begin{aligned}
\sum_{k=0}^{p} H\left(\sigma_k \begin{pmatrix} a & b \\ c & d\end{pmatrix}{\sigma_k}^{-1}\right) &= \left(\int_{\alpha\tau_o}^{i\infty} - \int_{\tau_o}^{i\infty}\right)\left(\sum_{k=0}^{p} f(\sigma_k z)\,d(\sigma_k z)\right) \\
&= a(p)\left(\int_{\alpha\tau_o}^{i\infty} - \int_{\tau_o}^{i\infty}\right) f(z)\,dz \\
&= a(p)H(\alpha)
\end{aligned}$$

by the properties of the p^{th} Hecke operator.

Property (5): $H\left(\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1\end{smallmatrix}\right)\right) = 0$.

Proof: By definition $H\left(\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1\end{smallmatrix}\right)\right) = F(\tau+1) - F(\tau)$. Since

$$F(\tau) = \sum_{n=1}^{\infty} \frac{a(n)}{n} e^{2\pi i n\tau}$$

which is periodic in τ we easily see that $F(\tau+1) = F(\tau)$.

§4. Equivalent forms of Szpiro's conjecture

We now give some equivalent forms of Szpiro's conjecture (2). Let

$$\Delta(z) = e^{2\pi iz} \prod_{n=1}^{\infty} \left(1 - e^{2\pi inz}\right)^{24}$$

be the Ramanujan cusp form of weight twelve for the full modular group. Then since the discriminant D of the elliptic curve

$$E: \; y^2 = x^3 - ax - b$$

can be expressed

$$D = \frac{\Delta\left(-\frac{\Omega_1}{\Omega_2}\right)}{2\pi^{12}{\Omega_2}^{12}}$$

and, without loss of generality we may assume that $|\frac{\Omega_1}{\Omega_2}| > 1$, we see that $|\Delta\left(-\frac{\Omega_1}{\Omega_2}\right)|$ is absolutely bounded from above by some fixed constant $c > 0$. It follows that we have $D < c/({\Omega_2}^{12})$. Hence, a lower bound of type

$$\Omega_2 \gg \frac{1}{N^\kappa} \tag{4}$$

for some fixed constant $\kappa > 0$ would give Szpiro's conjecture.

Now, since e_1, e_2, e_3, are roots of $4x^3 - ax - b = 0$ with a, b integers and $y^2 = 4x^3 - ax - b$ is a Frey curve, we easily see that $|e_i - e_j| \gg 1$ for $1 \le i < j \le 3$. Consequently the discriminant D satisfies $D \gg |e_i - e_j|^2$ for $i \neq j$. Hence

$$\begin{aligned}
\Omega_2 &= 2\int_{e_2}^{e_3} \frac{dx}{\sqrt{4(x-e_1)(x-e_2)(x-e_3)}} \\
&\geq \frac{1}{\sqrt{e_3 - e_1}\,(e_3 - e_2)} \int_{e_2}^{e_3} dx \\
&\geq \frac{1}{\sqrt{e_3 - e_1}} \\
&\gg D^{-\frac{1}{4}}.
\end{aligned}$$

So if Szpiro's conjecture is true, this yields a lower bound of type (4). Similarly for Ω_1. It follows that Szpiro's conjecture is equivalent to lower bounds of type (4) for the periods of E. If we assume the conjecture of Taniyama and Weil, then certain properties of the Shimura map $H : \Gamma_o(N) \to \Lambda$ as defined in §3 can be shown to be equivalent to Szpiro's conjecture. We have the following conjecture.

Conjecture(4); *Let $N \to \infty$. There exists a fixed constant $\kappa > 0$ such that if $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_o(N)$ with $|a|, |b|, |c|, |d| \le N^2$ then*

$$H\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = m\Omega_1 + n\Omega_2$$

with $|m|$, $|n| \ll N^\kappa$.

Assuming the Taniyama-Weil conjecture, it can be shown that conjecture (4) is equivalent to Szpiro's conjecture (1). Moreover, the assumption that $|a|, |b|, |c|, |d| \leq N^2$ can be replaced by the simpler assumption that $|c| \leq N^2$. This is because $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ is in the kernel of H which implies that we can always arrange $|a|, |b|, |d| \leq |c|$ after a suitable left or right multiplication by upper triangular matrices in $\Gamma_o(N)$. On the basis of numerical evidence, however, it seems we may take κ in conjecture (4) arbitrarily small as $(N \to \infty)$ if we restrict ourselves to matrices $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)$(satisfying $|c| \leq N^2$) which form a minimal set of generators for $\Gamma_o(N)$, but this seems hopelessly difficult to prove at the present time.

To prove Szpiro's conjecture, it suffices to assume the existence of a homomorphism $H : \Gamma_o(N) \to \Lambda$, satisfying properties (1) to (5), and in addition satisfying conjecture (4). In this context, conjecture (4) is a conjecture concerning a group homomorphism between a non-abelian group of rank $\approx N/6$, (namely, $\Gamma_o(N)$), and a free abelian group of rank 2. A matrix $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right) \in \Gamma_o(N)$ will be termed close to the identity if $|a|, |b|, |c|, |d|$ are small. Conjecture (4) says that if $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)$ is close to the identity, then its image under H is close to the origin in the lattice Λ, (implying that H has properties analogous to a continuous function). A proof of conjecture (4) should make strong use of property (5) (Hecke operators).

We now give a sketch of the proof of the equivalence of conjectures (1) and (4). Let

$$f(z) = \sum_1^\infty a(n) e^{2\pi i n z}$$

be the normalized Hecke newform of weight 2 associated to E. Then we have for $\alpha \in \Gamma_o(N)$

$$H(\alpha) = \sum_1^\infty \frac{a(n)}{n} \left[e^{2\pi i n \alpha(\tau)} - e^{2\pi i n \tau} \right],$$

which is independent of τ in the upper half plane. If we define

$$L_f(s, \theta) = \sum_1^\infty \frac{a(n)}{n^s} e^{2\pi i n \theta}$$

and

$$H_s(\alpha, \tau) = \sum_1^\infty \frac{a(n)}{n^{1+s}} \left[e^{2\pi i n \alpha(\tau)} - e^{2\pi i n \tau} \right],$$

then letting $\tau \to i\infty$ and $s \to 0$ it follows that

$$H(\alpha) = H_o(\alpha, i\infty) = L_f(1, \frac{a}{c}). \tag{5}$$

To obtain further information about $H(\alpha)$, we need the functional equation of $L_f(s, \frac{a}{c})$. This is obtained as follows. Let us put $z = -\frac{d}{c} + iy$, and $\alpha = \left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)$. Then we have $\alpha(z) = \frac{a}{c} + \frac{i}{c^2 y}$. It follows that

$$\begin{aligned} c^s \Gamma(s)\, L_f\left(s, -\frac{d}{c}\right) &= \int_0^\infty f\left(-\frac{d}{c} + iy\right) (cy)^s \,\frac{dy}{y} \\ &= \int_0^{\frac{1}{c}} f(z)(cy)^s \,\frac{dy}{y} \;+\; \int_{\frac{1}{c}}^\infty f(z)(cy)^s \,\frac{dy}{y} \qquad , \\ &= \int_1^\infty f\left(\frac{a+iy}{c}\right) y^{2-s} \,\frac{dy}{y} \;+\; \int_1^\infty f\left(\frac{-d+iy}{c}\right) y^s \,\frac{dy}{y} \end{aligned}$$

which gives the functional equation

$$c^s\,\Gamma(s)\,L_f\left(s,-\frac{d}{c}\right)=c^{2-s}\,\Gamma(2-s)\,L_f\left(2-s,\frac{a}{c}\right)$$

where $ad\equiv 1(c)$. The usual convexity argument then gives

$$\left|L_f\left(1,\frac{a}{c}\right)\right| \ll c^{\frac{1}{2}+\epsilon}. \tag{6}$$

For $\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right)\in\Gamma_o(N)$ with $|c|<N^2$, let us choose

$$\alpha=\begin{pmatrix} a & b\\ c & d\end{pmatrix}\begin{pmatrix} a & -b\\ -c & d\end{pmatrix}.$$

If $H(\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right))=m\Omega_1+n\Omega_2$, then by property (2), we have that

$$H(\alpha)=2n\Omega_2.$$

It then follows from this and equations (5), (6) that

$$|n\Omega_2|\ll N^{2+\epsilon}. \tag{7}$$

But if Szpiro's conjecture is true, then

$$|\Omega_2|\gg\frac{1}{N^\kappa}$$

for some $\kappa>0$. The inequality (6) yields

$$|n|\ll N^{\kappa+2+\epsilon}.$$

A similar argument also works for the m-component of H. So we have shown that Szpiro's conjecture implies conjecture (4).

To show that conjecture (4) implies conjecture (1) is more difficult. Let us define $\chi:\mathbb{Z}/q\mathbb{Z}\to\{\pm 1\}$ to be a real primitive Dirichlet character (mod q). Consider the twisted L-series

$$L_f(s,\chi)=\sum_1^\infty\frac{a(n)\chi(n)}{n^s}.$$

If

$$G(\chi)=\sum_{a=1}^q\chi(a)e^{2\pi i\frac{a}{q}}$$

denotes the Gauss sum, then by the standard argument

$$G(\chi)L_f(s,\chi)=\sum_{b=1}^q\chi(b)L_f\left(s,\ \frac{b}{q}\right).$$

For any two integers b, q satisfying $(q, N) = 1$, $0 < b < q$, and $(b, q) = 1$ we can always choose suitable integers a, c so that $\gamma = \binom{a\ b}{c\ q}$ lies in $\Gamma_o(N)$. We then have

$$\sum_{b=1}^{q} \chi(b) H_s(\gamma, \tau) = \sum_{b=1}^{q} \chi(b) \sum_{1}^{\infty} \frac{a(n)}{n^{1+s}} \left[e^{2\pi i n \gamma(\tau)} - e^{2\pi i n \tau} \right].$$

Letting $\tau \to 0$ and $s \to 0$, yields

$$\sum_{b=1}^{q} \chi(b) H(\gamma) = \sum_{b=1}^{q} \chi(b) H_o(\gamma, 0) = \sum_{b=1}^{q} \chi(b) L_f\left(1, \frac{b}{q}\right)$$

since

$$\sum_{b=1}^{q} \chi(b) = 0.$$

It follows that

$$\sum_{b=1}^{q} \chi(b) H\left(\begin{pmatrix} a & b \\ c & q \end{pmatrix}\right) = G(\chi) L_f(1, \chi). \tag{8}$$

If $\chi(-1) = -1$, so that χ is an odd character, then the substitution $b \mapsto -b$, $c \mapsto -c$ does not change the value of the left side of equation (8) since we can sum over any set of residues (mod q). But by property (2) of the homomorphism H this implies that $G(\chi)L_f(1, \chi)$ must be pure imaginary, and hence must be an integral multiple of the imaginary period Ω_2.

Now, by a theorem of Waldspurger [W], (see Kohnen [K]) it follows that $L_f(1, \chi)$ is the square of a Fourier coefficient of a cusp form of weight $\frac{3}{2}$. Applying the Rankin-Selberg method, as in Kohnen and Zagier's proof [K-Z] of the Goldfeld-Viola conjecture [G-V] on mean values of $L_f(1, \chi)$ one obtains

$$\sum_{q \ll N^2} L_f(1, \chi) \sim N^2.$$

Since $|G(\chi)| = \sqrt{q}$, it follows that for some twist χ with conductor $q \ll N^2$

$$G(\chi) L_f(1, \chi) \gg N.$$

Consequently, if we assume conjecture (4), there is an integer m satisfying $m \ll N^{2+\kappa}$ for some fixed $\kappa > 0$ such that

$$m\Omega_2 \gg N.$$

We then obtain that

$$\Omega_2 \gg N^{-1-\kappa},$$

and as shown earlier, this implies conjecture (1).

In conclusion, I should like to focus on yet another equivalence to Szpiro's conjecture (2). Kohnen [K] has shown that associated to a normalized newform $f(z)$

of weight 2 for $\Gamma_o(N)$ there is a cusp form $g(z)$ of weight $\frac{3}{2}$ for $\Gamma_o(4N)$ whose q^{th} Fourier coefficient $c(q)$ is given by

$$\frac{c(q)^2}{\langle g,g\rangle} = \frac{2^{\nu(N)}\sqrt{q}\;\; L_f(1,\chi)}{\pi\,\langle f,f\rangle} \tag{9}$$

where $\nu(N)$ denotes the number of prime factors of N and for cusp forms f_1, f_2 of weight $k \in \frac{1}{2}\mathbb{Z}$ for $\Gamma = \Gamma_o(M)$

$$\langle f_1, f_2\rangle = \frac{1}{[\Gamma_o(1):\Gamma]}\int_{\Gamma\backslash H} f_1(z)\overline{f_2(z)}y^k\,\frac{dx\,dy}{y^2}$$

denotes the Peterson inner product (Here H is the upper half plane).

Clearly, the left hand side of (9) is independent of the normalization of g. Let us normalize g so that $c(q) \in \mathbb{Z}$ for all q and

$$G(\chi)L_f(1,\chi) = c(q)^2\Omega_2.$$

Szpiro's conjecture is then equivalent to the bound

$$\langle g,g\rangle \ll N^c$$

for some fixed constant $c > 0$. This follows easily from (9) by the estimate

$$\frac{1}{[\Gamma_o(1):\Gamma_o(N)]} \ll \langle f,f\rangle \ll 1. \tag{10}$$

To prove (10) note that

$$\begin{aligned}\int_{\Gamma_o(N)\backslash H} |f(z)|^2\,dxdy &\geq \int_1^\infty \int_0^1 |f(z)|^2 dxdy \\ &= \int_1^\infty \sum_1^\infty |a(n)|^2 e^{-4\pi n y} dy \\ &= \sum_1^\infty \frac{a(n)^2 e^{-4\pi n}}{4\pi n} \\ &\gg 1\end{aligned}$$

since $a(1) = 1$.

On the other hand, if we let $d(n)$ denote the number of divisors of an integer n, then the Fourier coefficients of f at an arbitrary cusp (see [D]) are bounded by $\sqrt{n}d(n)$. It follows that

$$\begin{aligned}\int_{\Gamma_o(N)\backslash H} |f(z)|^2 dxdy &= \sum_{\gamma\in\Gamma_o(N)\backslash\Gamma_o(1)} \int_{\Gamma_o(1)\backslash H} |f(\gamma z)|^2\,\mathrm{Im}(\gamma z)^2\frac{dxdy}{y^2} \\ &\ll \sum_\gamma \int_{\frac{\sqrt{3}}{2}}^\infty \int_0^1 f(\gamma z)\,\mathrm{Im}(\gamma z)^2\frac{dxdy}{y^2} \\ &\ll [\Gamma_o(1):\Gamma_o(N)]\sum_{n=1}^\infty nd(n)^2 e^{-2\pi\sqrt{3}n} \\ &\ll [\Gamma_o(1):\Gamma_o(N)]\end{aligned}.$$

We have seen that for an integral weight modular form f with relatively prime rational integer Fourier coefficients , it is possible to give an absolute bound for $\langle f, f\rangle$ which is independent of the level. This is due to the fact that the n^{th} Fourier coefficient $a(n)$ is bounded by $\sqrt(n)d(n)$. If we knew that $|a(n)| \leq Cn^{\theta}$ for constants C, θ independent of n (but possibly C depending on N) then by the properties of the Hecke operators we would have

$$a(p) \approx 2a(p^M)^{\frac{1}{M}}$$

for rational primes p. Letting $M \to \infty$, it follows by a simple argument that $|a(n)| \leq d(n)n^{\theta}$; and in effect, the constant C drops out of the picture. In the half integral weight case, however, this does not happen because there are not enough Hecke operators.

BIBLIOGRAPHY

[**D**] Deligne, P. *La conjecture de Weil, I*, Publ. Math IHES, **43** (1974), 273-308.

[**F1**] Frey, G. *Rationale Punkte auf Fermatkurven und getwistete Modulkurven,* J. Reine. Angew. Math. **331** (1982), 185-191.

[**F2**] Frey, G. *Links between stable elliptic curves and certain diophantine equations,* Annales Universitatis Saraviensis, Vol 1, No. 1 (1986), 1-39.

[**F3**]Frey, G. *Links between elliptic curves and the solutions of the equation* $A-B = C$, preprint, Sarrebrüken RFA.

[**G-V**] Goldfeld, D. & Viola, C. *Mean values of L-functions associated to elliptic, Fermat and other curves at the centre of the critical strip,* J. Number Theory **11** (1979), 305-320.

[**H-S**] Hindry, M & Silverman, J. *The canonical height and elliptic curves,* preprint (1987).

[**K**] Kohnen, W. *Fourier coefficients of modular forms of half-integral weight,* Math. Ann. **271** (1985), 237-268.

[**K-Z**] Kohnen, W. & Zagier, D. *Values of L-series of modular forms at the center of the critical strip,* Invent. Math. **64** (1981), 175-198.

[**M1**] Mazur, B. *Modular curves and the Eisenstein ideal,* Inst. Hautes Études Sci. Publ. Math. **47** (1977), 33-186.

[**M2**] Mazur, B. Letter to J-F Mestre.

[**M-Sw**] Mazur, B. & Swinnerton-Dyer, P. *Arithmetic of Weil curves,* Inventiones Math. **25** (1974), 1-61.

[**Ost**] Osterlé , J. *Nouvelles approches du Théorème de Fermat,* Sem. Bourbaki, n^o 694 (1987-88), 694-01 - 694-21.

[**R**] Ribet, K. *Lectures on Serre's conjectures,* MSRI preprint (1987).

[**S1**] Serre, J-P. Lettre à J-F Mestre (13 Août 1985), To appear in Current trends in arithmetic.

[**S2**] Serre, J-P. *Sur les representations modulaires de degré 2 de* Gal$(\overline{Q}/Q)$, Duke Math. J. **54** (1987), 179-230.

[**Sh**] Shimura, G. *On the factors of the jacobian variety of a modular function field J*, Math. Soc. Japan **25** (1973), 523-544.
[**Szp1**] Szpiro, L. *Seminaire sur les pinceaux de courbes de genre au moins deux,* Astérisque, exposé n^o 3, **86** (1981), 44-78.

[**Szp2**] Szpiro, L. *Présentation de la théorie d'Arakélov,* Contemporary Math. **67** (1987), 279-293.

[**W**] Waldspurger, J-L. *Sur les coefficients de Fourier des formes modulaires de poides demi-entier,* J. Math. Pures Appl. **60** (1981), 375-484.

Dorian Goldfeld
Department of Mathematics
Columbia University
New York City,
New York
10027