UN PASEO POR LA MODULARIDAD

HECTOR PASTEN

1. Introducción

Se atribuye a Eichler la siguiente cita: *Solo hay cinco operaciones aritméticas elementales: suma, resta, multiplicación, división y formas modulares.* Este artículo está orientado principalmente a quienes no están familiarizados con la quinta. No es nuestra intención dar una introducción completa a la teoría ni mucho menos, sino más bien acompañar al lector en un breve paseo por el interesante mundo de las formas modulares. Esperamos que al concluir esta corta visita, el lector pueda tener una idea del sentido de la frase citada al comienzo.

Comenzaremos presentando a las protagonistas. En pocas palabras, una forma modular es una función $f(z)$ con *propiedades analíticas buenas* y que *se transforma de manera sencilla* bajo algunos cambios de variable que vamos a explicar más adelante. Estas funciones aparecen vinculadas a una gran cantidad de objetos aritmético-geométricos (por ejemplo, ecuaciones diofantinas, variedades algebraicas, etc.) y pueden ser usadas para resolver diversos problemas en teoría de números y otras áreas, con aplicaciones tan espectaculares como la demostración del Último Teorema de Fermat:

Si $n \geq 3$ entonces toda solución entera de la ecuación $x^n + y^n = z^n$ cumple $xyz = 0$.

Iniciaremos nuestra visita dando ejemplos de formas modulares y *después* daremos la definición junto con elementos básicos de la teoría (aunque parezca anti-pedagógico). Y por supuesto, nuestro paseo no estaría completo sin dar un vistazo a lo que es posiblemente la mayor atracción para los visitantes; explicando las ideas de como se utilizaron formas modulares en la demostración del Último Teorema de Fermat.

2. Formas modulares aparecen en el camino

Un teorema de Lagrange dice que todo entero $n \geq 0$ es suma de cuatro cuadrados de numeros enteros. Por ejemplo $2 = 1^2 + 1^2 + 0^2 + 0^2$ y $14 = 3^2 + 2^2 + 1^2 + 0^2$. Después de saber esto, naturalmente queremos saber de cuantas maneras. Definimos $r(n)$ como la cantidad de maneras de escribir n como suma de cuatro cuadrados de números enteros, tomando en cuenta el orden. Por ejemplo $r(0) = 1$ porque sólo se puede $0 = 0^2 + 0^2 + 0^2 + 0^2$, mientras que $r(1) = 8$ porque se puede $1^2 + 0 + 0 + 0$ con sus reordenamientos, y $(-1)^2 + 0 + 0 + 0$ con sus reordenamientos. Vamos a buscar una expresión conveniente para una función que guarde la información de todos los $r(n)$ a la vez. Notemos que

$$\left(\sum_{n\in\mathbb{Z}} q^{n^2}\right)^4 = \sum_{n_1,n_2,n_3,n_4\in\mathbb{Z}} q^{n_1^2+n_2^2+n_3^2+n_4^2} = \sum_{m=0}^{\infty} \sum_{n_1^2+n_2^2+n_3^2+n_4^2=m} q^m = \sum_{m=0}^{\infty} r(m)q^m.$$

Con esto en mente definimos

$$\theta = \sum_{n\in\mathbb{Z}} q^{n^2} = 1 + 2\sum_{n=1}^{\infty} q^{n^2}$$

y por lo tanto θ^4 es igual a $R = r(0) + r(1)q + r(2)q^2 + \cdots$, la función generadora de $r(n)$. Ahora queremos investigar propiedades de la función θ, para esto va a ser conveniente el cambio de variable $q = e^{2i\pi z}$ de modo que θ y R son series de Fourier en la variable compleja z

$$\theta(z) = 1 + 2\sum_{n=1}^{\infty} e^{2i\pi n^2 z}, \quad R(z) = \sum_{n=0}^{\infty} r(n)e^{2i\pi nz}.$$

Date: August 26, 2012.

El lector puede verificar fácilmente que estas series convergen cuando la parte imaginaria de z es estrictamente positiva, así que definen funciones holomorfas (diferenciable compleja) en el semi-plano superior $\mathfrak{h} = \{z \in \mathbb{C} : \Im(z) > 0\}$.

Naturalmente $\theta(z)$ es 1-periódica, es decir $\theta(z+1) = \theta(z)$, porque $e^{2i\pi n^2 z}$ lo es. Sin embargo, θ satisface otras reglas de transformación mucho menos evidentes al hacer otros cambios de variable, como por ejemplo

$$\theta\left(\frac{9z+2}{4z+1}\right) = \sqrt{4z+1}\theta(z).$$

Para poder enunciar el resultado general, definimos el conjunto de matrices

$$\Gamma_0(4) = \left\{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d \in \mathbb{Z}, \det\gamma = 1, c \equiv 0 \mod 4\right\}$$

donde $c \equiv 0 \mod 4$ significa que 4 divide c. Por ejemplo $\begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix} \in \Gamma_0(4)$. Dada $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4)$ definimos $\gamma z = \frac{az+b}{cz+d}$. Un ejercicio sencillo es verificar que, si $z \in \mathfrak{h}$ entonces $\gamma z \in \mathfrak{h}$ (ayuda: usar el hecho que $\det\gamma = 1$ al calcular la parte imaginaria de γz). Con esta notación, se tiene:

Teorema 2.1. *Si* $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4)$ *entonces para* $z \in \mathfrak{h}$

$$\theta(\gamma z) = w\sqrt{cz+d}\theta(z)$$

donde $w \in \{1, -1, i, -i\}$.

Notar que la elección de raiz cuadrada en $\sqrt{cz+d}$ no importa porque w es alguna raiz cuarta de 1, aunque se puede ser más preciso sobre el valor de w. De todas formas, tomando la cuarta potencia obtenemos la siguiente consecuencia:

Teorema 2.2. *Si* $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4)$ *entonces para* $z \in \mathfrak{h}$

$$R(\gamma z) = (cz+d)^2 R(z).$$

Tanto $\theta(z)$ como $R(z)$ son *formas modulares* para $\Gamma_0(4)$. El exponente de $cz+d$ en la regla de transformación se llama el *peso*; de esta manera θ tiene peso $1/2$ y $R(z)$ tiene peso 2.

Volviendo al tema de las sumas de cuatro cuadrados, ¿podemos obtener alguna información sobre $r(n)$ ahora que sabemos que R es una forma modular? es decir, ¿la regla de transformación es útil para estudiar $r(n)$? La respuesta es *sí*, esto será explicado más adelante.

Si las reglas de transformación de $\theta(z)$ indicadas arriba no son suficientes para entusiasmar al lector, aquí hay otra más no incluida en el teorema 2.1 (una transformación "extra"):

Proposición 2.3. *Si* $z \in \mathfrak{h}$ *entonces*

$$\theta\left(\frac{-1}{4z}\right) = \sqrt{-2iz}\theta(z)$$

donde la raiz cuadrada se elige tal que $\sqrt{1} = 1$.

Esta regla de transformación es más sencilla de demostrar que las del teorema 2.1. Consiste en una aplicación de la fórmula de sumación de Poisson, el lector interesado puede tratar de demostrarla por si mismo.

3. Fabricando formas modulares

Demostrar la regla de transformación de $\theta(z)$ dada por el teorema 2.1 es difícil. Lo que vamos a hacer ahora es construir funciones que tienen reglas de transformación similares, pero mucho más simples de verificar.

Primero, en lugar de usar $\Gamma_0(4)$ vamos a usar todas las matrices de 2×2 a coeficientes enteros y determinante 1

$$\mathrm{SL}_2(\mathbb{Z}) = \left\{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d \in \mathbb{Z}, \det\gamma = 1\right\}$$

y definimos γz con una fracción como antes. Notar que $\mathrm{SL}_2(\mathbb{Z})$ es más grande que $\Gamma_0(4)$, pero también es más sencillo. Las matrices de $\mathrm{SL}_2(\mathbb{Z})$ actúan en los vectores $(m,n) \in \mathbb{Z}^2$ con coordenadas enteras biyectivamente, ya que la matriz inversa de γ también está en $\mathrm{SL}_2(z)$ (tiene coeficientes enteros).

Si $k \geq 4$ es par, definimos la serie de Eisenstein

$$G_k(z) = \sum_{(m,n)\in\mathbb{Z}^2-(0,0)} \frac{1}{(mz+n)^k}.$$

La sumatoria es sobre todos los pares de enteros (m,n) excepto cuando $m = n = 0$. Como $k \geq 4$ esta serie converge absolutamente así que podemos reordenarla como queramos. Además, si k fuera impar entonces ¡la serie completa se cancela y daría cero! El siguiente teorema es un ejercicio sencillo.

Teorema 3.1. *Si* $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ *entonces*

$$G_k(\gamma z) = (cz+d)^k G(z).$$

La demostración se obtiene a partir de

$$\frac{1}{(m\gamma z+n)^k} = (cz+d)^k \frac{1}{((ma+nc)z+(mb+nd))^k}$$

y de la observación que γ aplicada (como matriz) al vector (m,n) es igual a $(ma+nc, mb+nd)$ de manera que la sumatoria que define $G_k(z)$ solo se reordena. Si aplicamos este teorema al caso particular $\gamma = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ obtenemos que $G_k(z+1) = G_k(z)$ por lo tanto podemos esperar que G_k se puede expresar como serie de Fourier. Esto es correcto y los coeficientes de Fourier se pueden dar explícitamente. Para nuestra sorpresa ¡estos coeficientes codifican información aritmética! Primero necesitamos introducir notación.

Definimos $\sigma_r(n) = \sum_{d|n} d^r$ donde la suma es sobre los divisores positivos de n. Por ejemplo $\sigma_2(6) = 1^2 + 2^2 + 3^2 + 6^2 = 50$. La función zeta de Riemann es definida por $\zeta(s) = \sum_{n\geq 1} n^{-s}$ y esta serie converge para $\Re(s) > 1$. Por ejemplo, es sabido que

$$\zeta(2) = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots = \frac{\pi^2}{6}.$$

Teorema 3.2. *Para* $k \geq 4$ *par, consideramos la serie de Eisenstein* $G_k(z)$*. Si* $z \in \mathfrak{h}$ *entonces*

$$G_k(z) = 2\zeta(k) + \frac{2(2i\pi)^k}{(k-1)!} \sum_{n=1}^{\infty} \sigma_{k-1}(n) e^{2i\pi nz}.$$

La función $G_k(z)$ es una forma modular para $\mathrm{SL}_2(\mathbb{Z})$ de peso k, porque su regla de transformación tiene el factor $(cz+d)^k$.

4. Un poco de teoría

Ahora que tenemos ejemplos en mente vamos a dar la definición de una forma modular en el caso más sencillo: formas modulares para $\mathrm{SL}_2(\mathbb{Z})$ de peso entero. Esto excluye por ejemplo la función $\theta(z)$ porque tiene peso $1/2$ (no entero) y sólo es modular en $\Gamma_0(4)$ (no en todo $\mathrm{SL}_2(\mathbb{Z})$).

Sea $k \geq 0$ un entero. Una *forma modular* de peso k para $\mathrm{SL}_2(\mathbb{Z})$ es una función $f : \mathfrak{h} \to \mathbb{C}$ con las siguientes propiedades:

- (condición analítica) f es diferenciable compleja y más aun, se puede expresar como serie de Fourier sin índices negativos:

$$f(z) = \sum_{n=0}^{\infty} c_n e^{2ni\pi z}.$$

- (condición modular) Si $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ entonces

$$f(\gamma z) = (cz+d)^k f(z).$$

La definición general con otros conjuntos de matrices como $\Gamma_0(4)$ y no sólo $\mathrm{SL}_2(\mathbb{Z})$ es un poco más técnica e involucra la noción de "ser regular en las cúspides", no lo vamos a explicar aquí. Por otro lado, si queremos hablar de pesos no enteros (como el caso de $\theta(z)$) entonces hay que modificar la condición de modularidad cambiando el factor $(cz+d)^k$ por $w(cz+d)^k$ donde w es cierta raiz de 1 (similar al teorema 2.1), preferimos evitar detalles técnicos complicados.

Hay otros conjuntos de matrices interesantes, como por ejemplo

$$\Gamma_0(N) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \gamma \in \mathrm{SL}_2(\mathbb{Z}), c \equiv 0 \mod N \right\}.$$

En particular con $N=4$ obtenemos exactamente el conjunto $\Gamma_0(4)$ definido antes, y con $N=1$ obtenemos $\Gamma_0(1) = \mathrm{SL}_2(\mathbb{Z})$. Una forma modular para $\Gamma_0(N)$ se dice[1] que tiene *nivel* N.

En general, el conjunto de todas las formas modulares de nivel N y peso k se denota por $M_k(\Gamma_0(N))$. Un ejercicio sencillo que dejamos al lector es que $M_k(\Gamma_0(N))$ es un espacio vectorial complejo. El siguiente resultado es fundamental:

Teorema 4.1. *Cada uno de los espacios $M_k(\Gamma_0(N))$ tiene dimensión finita. Si $k<0$ estos espacios son nulos, y si $k \geq 0$ entonces*

$$\dim_{\mathbb{C}} M_k(\Gamma_0(N)) \leq 1 + \frac{kN}{12} \prod_{p|N} \left(1 + \frac{1}{p}\right).$$

Calculemos unos ejemplos: con $k=4$, $N=1$ tenemos $\dim_{\mathbb{C}} M_4(\mathrm{SL}_2(\mathbb{Z})) \leq 1 + 1/3$ pero $G_4 \in M_4(\mathrm{SL}_2(\mathbb{Z}))$ así que $M_4(\mathrm{SL}_2(\mathbb{Z}))$ tiene dimensión 1 y es generado por G_4. Con $k=2$, $N=4$ tenemos

$$\dim_{\mathbb{C}} M_2(\Gamma_0(4)) \leq 1 + \frac{2 \cdot 4}{12}(1 + 1/2) = 2$$

esta desigualdad va a ser útil en la sección siguiente.

El espacio $M_k(\Gamma_0(N))$ tiene un subespacio $S_k(\Gamma_0(N))$ muy interesante llamado *espacio cuspidal*. Consiste de todas las formas modulares $f \in M_k(\Gamma_0(N))$ que satisfacen la siguiente condición técnica adicional: "f se anula en las cúspides". No vamos a entrar en detalles sobre el significado de esto, pero podemos enunciar una consecuencia:

Proposición 4.2. *Si $f \in S_k(\Gamma_0(N))$ entonces el desarrollo de Fourier de f no tiene término constante. Es decir $f(z) = \sum_{n=1}^{\infty} c_n q^n$ no tiene el término c_0 (donde $q = e^{2i\pi z}$).*

Veamos algunos ejemplos en el caso de peso $k=2$. Primero, tenemos $S_2(\Gamma_0(N)) = 0$ para $N = 1, 2, \ldots, 10$. El primer ejemplo de espacio cuspidal de peso 2 que no es nulo corresponde a $S_2(\Gamma_0(11))$, el cual tiene dimensión 1 y es generado por cierta forma modular cuya expansión de Fourier comienza así:

$$q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 + \ldots$$

donde $q = e^{2i\pi z}$ como siempre. Un ejemplo con dimensión más grande es $S_2(\Gamma_0(64))$ cuya dimensión es 3 y tiene la siguiente base (sólo damos los desarrollos de Fourier hasta q^{26}):

$$f_1 = q + 2q^5 - 3q^9 - 6q^{13} + 2q^{17} - q^{25} + \cdots$$
$$f_2 = q - 2q^5 - 3q^9 + 6q^{13} + 2q^{17} - q^{25} + \cdots$$
$$f_3 = q^2 - 2q^{10} - 3q^{18} + 6q^{26} + \cdots$$

Estos desarrollos de Fourier fueron calculados con el programa *Sage*, disponible en `www.sagenb.org` de forma gratuita y para uso online (no es necesario descargarlo ni instalarlo, aunque se puede). Es un programa orientado a matemáticas; es potente, fácil de usar y existen varios tutoriales online.

[1] La noción general de nivel requiere hablar de otros conjuntos de matrices llamados $\Gamma(N)$, pero para simplificar la exposición nos restringimos a trabajar sólo con $\Gamma_0(N)$.

5. De vuelta a las sumas de cuadrados

Volvamos al problema de las sumas de cuadrados estudiando $r(n)$. Sabemos que $R(z)$ es una forma modular de peso 2 para $\Gamma_0(4)$, es decir $R(z) \in M_2(\Gamma_0(4))$. Además sabemos que $M_2(\Gamma_0(4))$ es un espacio vectorial de dimensión finita. Con esta información, el plan es el siguiente: encontrar una base de $M_2(\Gamma_0(4))$ con formas modulares sencillas, expresar $R(z)$ como combinación lineal de esa base, y deducir una fórmula para $r(n)$ mirando los coeficientes de Fourier.

Hasta ahora, las formas modulares más sencillas que hemos encontrado son las series de Eisenstein G_k pero lamentablemente ellas sólo son formas modulares de peso $k \geq 4$ par, mientras que $R(z)$ tiene peso 2. Aun así, si copiamos la serie de Fourier de G_k y simplemente la escribimos para $k = 2$ podemos definir

$$G_2(z) := 2\zeta(2) + \frac{2(2i\pi)^2}{(2-1)!}\sum_{n=1}^{\infty}\sigma_{2-1}(n)e^{2i\pi nz} = \frac{\pi^2}{3} - 8\pi^2\sum_{n=1}^{\infty}\sigma_1(n)q^n$$

donde hemos usado el hecho que $\zeta(2) = \pi^2/6$ y la notación $q = e^{2i\pi z}$. La verdad es que G_2 define una función en $\mathfrak{h}$ pero no es una forma modular, al menos no en el sentido que definimos antes. El problema es que para demostrar la regla de transformación de G_k ($k \geq 4$) reordenábamos series infinitas lo que se justificaba por la convergencia absoluta, pero para $k = 2$ no se cumplía convergencia absoluta, sólo convergencia condicional. Peor todavía: el espacio $M_2(\mathrm{SL}_2(\mathbb{Z}))$ es cero, así que es seguro que G_2 no es una forma modular de peso 2 para $\mathrm{SL}_2(\mathbb{Z})$.

Lo que ocurre es que la regla de transformación de G_2 también es sencilla, pero es distinta.

Teorema 5.1. *Se tiene que* $G_2(z+1) = G_2(z)$ *y* $G_2(-1/z) = z^2G_2(z) - 2i\pi z$.

Las transformaciones dadas aquí sólo corresponden a las matrices $t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ y $s = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, pero se puede demostrar que toda matriz de $\mathrm{SL}_2(\mathbb{Z})$ es un producto de potencias de s y t por lo tanto la información del teorema basta. Uno dice que G_2 es una forma *quasi-modular de peso* 2 para $\mathrm{SL}_2(\mathbb{Z})$ porque su regla de transformación está muy cerca de ser modular (sólo ese $2i\pi z$ está estorbando) y además el factor z^2 tiene exponente 2.

Aun así, podemos usar G_2 para producir formas modulares de peso 2 con el siguiente truco:

Primero, para simplificar la historia definimos

$$E_2(z) := \frac{3}{\pi^2}G_2(z) = 1 - 24\sum_{n=1}^{\infty}\sigma_1(n)q^n.$$

El lector debe apreciar la simplicidad de la fórmula de $E_2(z)$, esto va a ser útil.

Entonces podemos definir $E_{2,2}(z) = E_2(z) - 2E_2(2z)$ y $E_{2,4}(z) = E_2(z) - 4E_2(4z)$. El efecto de restar una modificación de E_2 es que ahora el término $2i\pi z$ que estorbaba en la regla de transformación de G_2 se va a cancelar, pero hay un precio que pagar: $E_{2,2}(z)$ y $E_{2,4}(z)$ sólo son formas modulares de peso 2 para $\Gamma_0(4)$, no para el conjunto $\mathrm{SL}_2(\mathbb{Z})$ completo. Este precio es aceptable, ya que nosotros estamos interesados en $R(z)$ que vive en $M_2(\Gamma_0(4))$.

Por lo tanto hemos producido dos formas modulares nuevas en $M_2(\Gamma_0(4))$ que son

$$\begin{aligned} E_{2,2}(z) &= 1 - 24\sum_{n=1}^{\infty}\sigma_1(n)q^n - 2\left(1 - 24\sum_{n=1}^{\infty}\sigma_1(n)q^{2n}\right) \\ &= -1 - 24\sigma_1(1)q - 24(\sigma_1(2) - 2\sigma_1(1))q^2 - 24\sigma_1(3)q^3 + \cdots \\ &= -1 - 24q - 24q^2 - 96q^3 + \cdots \end{aligned}$$

y (calculando similarmente)

$$E_{2,4}(z) = -3 - 24q - 72q^2 - 96q^3 + \cdots.$$

Mirando los primeros coeficientes de Fourier es evidente que $E_{2,2}(z)$ y $E_{2,4}(z)$ son linealmente independientes. Más aun, sabemos (por la seccion anterior) que la dimensión de $M_2(\Gamma_0(4))$ es a lo más 2 por lo tanto obtenemos que $E_{2,2}(z)$ y $E_{2,4}(z)$ son una base de $M_2(\Gamma_0(4))$. Por lo tanto, existen α, β tales que $R(z) = \alpha E_{2,2}(z) + \beta E_{2,4}(z)$.

Recordemos que $R(z) = r(0) + r(1)q + \cdots = 1 + 8q + \cdots$ así que mirando los primeros coeficientes de Fourier tenemos:

$$\begin{aligned} R = \alpha E_{2,2} + \beta E_{2,4} &\Rightarrow 1 + 8q + \cdots = \alpha\left(-1 - 24q + \cdots\right) + \beta\left(-3 - 24q + \cdots\right) \\ &\Rightarrow 1 + 8q + \cdots = (-\alpha - 3\beta) + (-24\alpha - 24\beta)q + \cdots \\ &\Rightarrow \begin{cases} 1 = -\alpha - 3\beta \\ 8 = -24\alpha - 24\beta \end{cases} \end{aligned}$$

y así obtenemos $\alpha = 0$, $\beta = -1/3$. Por lo tanto

$$\begin{aligned} R(z) = -\frac{1}{3}E_{2,4}(z) &= \frac{-1}{3}\left(1 - 24\sum_{n=1}^{\infty}\sigma_1(n)q^n\right) + \frac{4}{3}\left(1 - 24\sum_{n=1}^{\infty}\sigma_1(n)q^{4n}\right) \\ &= 1 + 8\sum_{n=1}^{\infty}\left(\sigma_1(n) - 4\sigma_1(n/4)\right)q^n \end{aligned}$$

con la notación $\sigma_1(n/4) = 0$ cuando $n/4$ no es entero. Recordando que $R(z) = \sum_{n\geq 0} r(n)q^n$ obtenemos:

Teorema 5.2. *Para $n = 0$ tenemos $r(0) = 1$. Si $n > 0$ no es divisible por 4 entonces $r(n) = 8\sigma_1(n)$, y si $n > 0$ es divisible por 4 entonces $r(n) = 8(\sigma_1(n) - 4\sigma_1(n/4))$.*

Es decir ¡tenemos una fórmula explícita y sencilla para saber de cuántas formas un número $n \geq 0$ es suma de cuatro cuadrados enteros! Por ejemplo, ahora fácilmente sabemos que 6 se puede escribir como suma de cuatro cuadrados enteros en $r(6) = 8\sigma_1(6) = 8(1+2+3+6) = 96$ maneras (básicamente $6 = 2^2+1^2+1^2+0^2$ y todos los reordenamientos y cambios de signo dentro de los cuadrados no nulos).

La moraleja de esta historia es que si tenemos una secuencia $(a_n)_n$ y resulta que $\sum a_n q^n$ es una forma modular entonces la teoría de formas modulares nos permite obtener mucha información, incluso una fórmula exacta si andamos con suerte.

6. Funciones L

Digamos que tenemos una secuencia de números $a_1, a_2, \ldots$ que nos gustaría estudiar. Una práctica común en teoría de números y geometría es considerar la función generadora

$$F(s) = a_1 + \frac{a_2}{2^s} + \frac{a_3}{3^s} + \cdots = \sum_{n\geq 1}\frac{a_n}{n^s}.$$

Cuando la secuencia $(a_n)_n$, o equivalentemente la función $F(s)$, tiene "orígenes nobles" (digamos un problema aritmético o geométrico) entonces uno suele decir[2] que F es una *función L*.

Por ejemplo, la secuencia constante $1, 1, \ldots$ da origen a la función zeta de Riemann

$$\zeta(s) = \sum_{n=1}^{\infty}\frac{1}{n^s}$$

que aparece naturalmente al estudiar la distribución de los números primos, lo cual es un origen suficientemente noble así que este es nuestro primer ejemplo de una función L. Explicaremos brevemente cual es la relación con números primos: Euler demostró que $\zeta(s)$ se puede expresar como un producto sobre los primos p

$$\prod_p \frac{1}{1-p^{-s}} = \prod_p\left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots\right) = \sum_{n\geq 1}\frac{1}{n^s} = \zeta(s)$$

donde la primera igualdad se obtiene de sumar la serie geométrica, y la segunda igualdad es por la propiedad de factorización única de los enteros. De inmediato obtenemos aplicaciones en el estudio de números primos; por ejemplo tomando el límite cuando $s \to 1^+$ concluimos que existen infinitos primos pues la serie armónica $\sum_{n\geq 1} 1/n$ diverge.

Un producto de este tipo para una función L se conoce como *producto de Euler* y es una de las buenas propiedades que uno espera que las funciones L deberían cumplir.

[2]La primera persona en utilizar el nombre "función L" para este tipo de funciones fue Dirichlet en sus investigaciones sobre números primos en progresiones aritméticas. Nadie sabe con certeza por que eligió la letra L.

La función $\zeta(s)$ está muy relacionada a las formas modulares. Un cálculo sencillo nos muestra que

$$\zeta(s)\zeta(s-r) = \left(\sum_{a=1}^{\infty} \frac{1}{a^s}\right)\left(\sum_{b=1}^{\infty} \frac{b^r}{b^s}\right) = \sum_{n=1}^{\infty}\sum_{ab=n} \frac{b^r}{n^s} = \sum_{n=1}^{\infty} \frac{\sigma_r(n)}{n^s}$$

por lo tanto si $k \geq 4$ es par entonces $\zeta(s)\zeta(s-k+1)$ es básicamente (salvo un factor escalar) la función L asociada a los coeficientes de Fourier de la forma modular $G_k(z)$ (quasi-modular en el caso $k = 2$, lo cual es también muy bueno). Esto se puede enunciar de la siguiente forma:

Teorema 6.1. *Para $k \geq 2$ par, la función L definida por $\zeta(s)\zeta(s-k+1)$ es modular.*

El adjetivo "modular" sólo significa que la función L que estamos estudiando se puede recuperar a partir de cierta forma modular (o una forma quasi-modular, o alguna otra generalización del concepto de forma modular). Estamos frente a nuestro primer ejemplo de funciones L que son modulares.

Naturalmente uno se pregunta si la función $\zeta(s)$ es modular o no, en algún sentido aceptable. La respuesta es que *sí*, en el siguiente sentido:

Teorema 6.2. *Se tiene que*

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \int_0^{\infty} \frac{\theta(it/2)-1}{2} t^{s/2}\frac{dt}{t}$$

donde $\Gamma(s)$ es la función Gamma. En particular, la función $\zeta(s)$ es modular, asociada a la forma modular $\theta(z)$.

La demostración es un cálculo sencillo que detallamos a continuación (hay que hacer la sustitución $y = \pi n^2 t$ de la primera a la segunda línea):

$$\begin{aligned}\int_0^{\infty} (\theta(it/2)-1)t^{s/2}\frac{dt}{t} &= 2\int_0^{\infty}\left(\sum_{n=1}^{\infty} e^{-\pi n^2 t}\right) t^{s/2}\frac{dt}{t} = 2\sum_{n=1}^{\infty}\int_0^{\infty} e^{-\pi n^2 t} t^{s/2}\frac{dt}{t} \\ &= 2\sum_{n=1}^{\infty}\int_0^{\infty} e^{-y}\left(\frac{y}{\pi n^2}\right)^{s/2}\frac{dy}{y} = 2\pi^{-s/2}\sum_{n=1}^{\infty}\frac{1}{n^s}\int_0^{\infty} e^{-y}y^{s/2}\frac{dy}{y} \\ &= 2\pi^{-s/2}\zeta(s)\Gamma(s/2).\end{aligned}$$

Que una función L sea modular no es sólo interesante por el simple hecho de conectar ideas, sino que también hay consecuencias concretas. Por ejemplo, como la función $\zeta(s)$ es modular podemos usar las transformaciones de la función $\theta(z)$ para obtener información de $\zeta(s)$. Más precisamente, la regla de transformación "extra" de la función $\theta(z)$ evaluada en $z = it/2 \in \mathfrak{h}$ nos da $\theta(it^{-1}/2) = \sqrt{t}\theta(it/2)$ y así

$$\begin{aligned}2\pi^{-s/2}\Gamma(s/2)\zeta(s) &= \int_1^{\infty}(\theta(it/2)-1)t^{s/2}\frac{dt}{t} + \int_0^{1}(\theta(it/2)-1)t^{s/2}\frac{dt}{t} \\ &= \int_1^{\infty}(\theta(it/2)-1)t^{s/2}\frac{dt}{t} + \int_1^{\infty}(\theta(it^{-1}/2)-1)t^{-s/2}\frac{dt}{t} \\ &= \int_1^{\infty}(\theta(it/2)-1)t^{s/2}\frac{dt}{t} + \int_1^{\infty}(\sqrt{t}\theta(it/2)-1)t^{-s/2}\frac{dt}{t} \\ &= \int_1^{\infty}(\theta(it/2)-1)t^{s/2}\frac{dt}{t} + \int_1^{\infty}(\theta(it/2)-1)t^{(1-s)/2}\frac{dt}{t} + \int_1^{\infty}(t^{(1-s)/2}-t^{-s/2})\frac{dt}{t} \\ &= \int_1^{\infty}(\theta(it/2)-1)\left(t^{s/2}+t^{(1-s)/2}\right)\frac{dt}{t} - \frac{2}{(1-s)s}.\end{aligned}$$

Estos cálculos son correctos cuando $\Re(s) > 1$ (de hecho, $\zeta(s)$ sólo ha sido definida para $\Re(s) > 1$) pero el resultado final es una función definida en todo $\mathbb{C}$ porque el factor $(\theta(it/2)-1)$ decae exponencialmente cuando $t \to \infty$. Más aun, el resultado final es invariante bajo el cambio de variable $s \mapsto 1-s$. Con esta nueva información y usando propiedades conocidas de la función $\Gamma(s)$ obtenemos el siguiente teorema de Riemann:

Teorema 6.3. *La función $\zeta(s)$ se extiende a una función meromorfa en todo el plano complejo con un único polo en $s=1$ el cual es simple con residuo 1. Además, la función $\zeta(s)$ satisface la siguiente ecuación funcional:*

$$\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s)=\pi^{-(1-s)/2}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s).$$

La moraleja de esta historia es que, cuando sabemos que una función L es modular ganamos un montón de información útil (como por ejemplo una ecuación funcional). Cabe preguntarse entonces cuáles funciones L son modulares.

7. Algunas palabras sobre curvas elípticas

Una *curva elíptica* E es una ecuación de la forma $y^2=x^3+ax^2+bx+c$ donde $a,b,c\in\mathbb{Z}$ y las tres raíces de x^3+ax^2+bx+c son distintas. Por ejemplo la ecuación $y^2=x^3-10x$ es una curva elíptica, pero la ecuación $y^2=x^3-6x^2$ no lo es (se dice que es *singular*).

Las curvas elípticas son interesantes porque sus soluciones (x,y) forman un *grupo abeliano*; se pueden sumar geométricamente (aunque es necesario agregar un punto extra, el punto al infinito). Sin embargo, nosotros nos vamos a concentrar en su aritmética módulo primos p y no en la estructura de grupo.

Uno puede reducir una curva elíptica módulo un primo p pero no necesariamente el polinomio x^3+ax^2+bx+c va a seguir teniendo sus tres raíces distintas. Por ejemplo $y=x^3-10x$ módulo 5 se convierte en $y^2=x^3$, pero x^3 tiene sus tres raíces iguales así que $y=x^3-10x$ es singular módulo 5.

Dada una curva elíptica E, si al reducirla módulo p queda singular entonces decimos que E tiene *mala* reducción en p, de lo contrario decimos que tiene *buena* reducción en p. Obviamente una curva elíptica tiene mala reducción a lo más en un número finito de primos.

En realidad, dada una curva elíptica E uno puede hacer cambios de variable "admisibles" que permiten rescatar algunos primos de mala reducción y extender la lista de primos de buena reducción, pero no vamos a entrar en detalles técnicos.

Ahora vamos a describir dos números importantes asociados a una curva elíptica $E: y^2=x^3+ax^2+bx+c$, estos son el discriminante y el conductor.

El *discriminante* de E es definido[3] por $\Delta_E=16((\alpha-\beta)(\beta-\gamma)(\gamma-\alpha))^2$ donde α,β,γ son las raíces de x^3+ax^2+bx+c. Se puede demostrar que siempre $\Delta_E\in\mathbb{Z}$.

El *conductor* de E es un número que mide la mala reducción de E. Es de la forma

$$N_E=\prod_{p\text{ malo}}p^{e_p}$$

donde e_p mide el tipo de reducción mala en p (típicamente e_p es 1 o 2, excepto si $p=2,3$ donde puede tomar más valores). La receta para calcular el conductor es complicada (pero existe).

Ejemplo. Consideremos la curva elíptica $E: y^2=x^3-4x$. El polinomio x^3-4x tiene raíces $-2,0,2$ y estos números son distintos módulo p para cualquier primo $p>2$, y son iguales módulo $p=2$. Entonces el único primo de mala reducción es 2. Además el discriminante es $\Delta_E=2^{12}$ y se puede calcular que el conductor es $N_E=2^6=64$.

Cuando reducimos una curva elíptica módulo p queremos calcular $n_E(p)$ su cantidad de soluciones (x,y) módulo p. Hay un teorema de Hasse que nos da una buena aproximación del número de estas soluciones:

Teorema 7.1. *Si E es una curva elíptica con buena reducción en p entonces $n_E(p)=p-a_E(p)$ donde $|a_E(p)|<2\sqrt{p}$.*

Por ejemplo, si $E: y^2=x^3-4x$ como en el ejemplo anterior vamos a calcular $n_E(5)$. La lista de todas las soluciones módulo 5 es $(x,y)=(0,0),(2,0),(3,0)$ lo cual arroja $n_E(5)=3$ y por lo tanto $a_E(5)=2$. Vale la pena notar que $|a_E(5)|=2<2\sqrt{5}$ tal como dice el teorema de Hasse.

Los números $n_E(p)$ y $a_E(p)$ nos dan la misma información así que vamos a trabajar sólo con $a_E(p)$ que es más pequeño (en efecto, $n_E(p)\approx p$ mientras que $|a_E(p)|<2\sqrt{p}$).

[3]Distintos autores difieren en esta definición por una potencia de 2 chica, dependiendo de cual es el uso que se le va a dar a Δ_E

La secuencia $a_E(p)$ guarda información aritmética de la curva elíptica así que no es mala idea guardar toda esta información en un solo objeto. Así, definimos la función L de la curva elíptica E como el siguiente producto infinito

$$L(E,s) = (*) \times \prod_{p \text{ bueno}} \frac{1}{1 - a_E(p)p^{-s} + p^{1-2s}}$$

donde el factor $(*)$ es un producto finito que involucra sólo los primos de mala reducción que dividen con exponente 1 el conductor N_E. En particular, si ningún primo divide con exponente exactamente 1 al conductor entonces $(*) = 1$.

Este producto de Euler se puede expandir para obtener una suma infinita de la forma

$$L(E,s) = \sum_{n \geq 1} \frac{a'_E(n)}{n^s}$$

donde los números $a'_E(n)$ tienen las siguientes propiedades sencillas: $a'_E(1) = 1$, todos los $a'_E(n)$ son enteros y si p es un primo bueno entonces $a'_E(p) = a_E(p)$. En vista de esto último, vamos a anotar $a_E(n)$ en lugar de $a'_E(n)$ ya que no hay ambigüedad cuando $n = p$ es un primo de buena reducción.

Por ejemplo, si E es la curva elíptica $y^2 = x^3 - 4x$ como antes, entonces $a_E(5) = 2$. En este ejemplo, la función $L(E,s)$ no necesita el factor $(*)$ porque el único primo malo es 2, el conductor es $N_E = 2^6$ y por lo tanto ningún primo de mala reducción divide el conductor con exponente 1. Así que no es necesario agregar más factores en nuestra función L. Algunos valores adicionales de los $a_E(n)$ están dados en la siguiente tabla:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$a_E(n)$	1	0	0	0	2	0	0	0	-3	0	0	0	-6	0	0	0	2	0	0

por lo tanto

$$L(E,s) = 1 + \frac{2}{5^s} - \frac{3}{9^s} - \frac{6}{13^s} + \frac{2}{17^s} \cdots$$

(acaba de ocurrir un fenómeno curioso, ¿podrá descubrirlo el atento lector?).

Hacemos un comentario sobre la definición de la función L de una curva elíptica. Los factores en el producto infinito de la definición de $L(E,s)$ se ven misteriosos pero tienen una explicación y un contexto donde aparecen naturalmente: básicamente corresponden a la función zeta de la curva elíptica $E \mod p$ en el contexto de las llamadas Conjeturas de Weil. El lector interesado puede consultar el Apéndice C de [2] para aprender más acerca de las conjeturas de Weil (que ahora son teoremas). También aparecen naturalmente al estudiar las representaciones de Galois asociadas a E.

En resumen, partimos de una curva elíptica E que es un cierto tipo de ecuación con coeficientes enteros, estudiamos su aritmética reduciéndola módulo p y juntamos toda esa información para construir una función L de la forma $L(E,s) = \sum_{n \geq 1} a_E(n)n^{-s}$ donde los $a_E(n)$ son enteros y $a_E(1) = 1$.

8. La conjetura de Taniyama-Shimura

A finales de los años '50, Taniyama propuso el problema de si la función $L(E,s)$ estaba relacionada a las formas modulares. Esta pregunta fue hecha más precisa por Shimura, dando lugar a la conjetura siguiente

Todas las curvas elípticas son modulares.

Una afirmación tan fuerte no ganó mucho apoyo en la comunidad matemática desde el comienzo. Pero poco a poco se descubrió más información y se logró verificar algunos ejemplos que sirvieron de evidencia. Una forma mucho más precisa de la conjetura es la siguiente:

Sea E una curva elíptica de conductor N_E con función L dada por $L(E,s) = \sum_{n \geq 1} a_E(n)n^{-s}$. Entonces los números $a_E(n)$ son los coeficientes de Fourier de una forma modular f_E de peso 2 y nivel N_E que pertenece al espacio cuspidal. Es decir

$$f_E = a_E(1)q + a_E(2)q^2 + a_E(3)q^3 + \cdots \in S_2(\Gamma_0(N_E)).$$

Esta conjetura está demostrada gracias a las espectaculares ideas de Wiles. En 1993 Wiles hizo pública una demostración para el caso que N_E tiene todos sus factores primos con exponente 1 (el

caso *semi-estable*) pero la demostración tenía un error. Taylor asistió a Wiles para lograr superar las dificultades técnicas y en 1995 Wiles y Taylor publicaron una demostración correcta para el caso semi-estable (ver [10, 11]). Alrededor del año 2000, una serie de publicaciones por otros matemáticos lograron demostrar todos los casos restantes (no sólo el semi-estable) y estos trabajos se basaron en las ideas de Wiles.

Como consecuencia, la función $L(E,s)$ se extiende a todo el plano complejo y satisface una ecuación funcional similar a la de la función $\zeta(s)$:

$$N_E^{s/2}(2\pi)^{-s}\Gamma(s)L(E,s) = \pm N_E^{(2-s)/2}(2\pi)^{-(2-s)}\Gamma(2-s)L(E,2-s).$$

La demostración se basa en cálculos parecidos a los hechos en la sección 6 para la función $\zeta(s)$, pero usando un poco más de teoría.

9. Ciertamente, no cabía en el margen de una página

En el margen de una página del libro *Arithmetica* de Diofanto, Fermat escribió una vez que ningún cubo es suma de dos cubos[4], y ninguna potencia cuarta es suma de dos potencias cuartas, etc., y que él había encontrado una demostración maravillosa pero que lamentablemente no cabía en el margen de esa página. Esta demostración maravillosa nunca fue encontrada entre los documentos de Fermat, y el enunciado en cuestión pasó a la historia como el *último teorema de Fermat*, a la espera de una demostración. En lenguaje moderno, el último teorema de Fermat dice lo siguiente:

Si $n > 2$ entonces $x^n + y^n = z^n$ no tiene solución con enteros x, y, z estrictamente positivos.

El caso de potencias cuartas fue demostrado por Fermat, así que basta considerar el caso de potencias ℓ-ésimas con $\ell > 2$ un primo (todo exponente $n > 2$ es múltiplo de 4 o de algún primo impar). Como ℓ es impar, podemos escribir el problema en una forma más simétrica:

Si $\ell > 2$ es primo entonces toda solución entera de $x^\ell + y^\ell + z^\ell = 0$ cumple que x, y o z es cero.

Más aun, los casos siguientes (entre otros) son sabidos: $\ell = 3$ (Euler), $\ell = 5$ (Legendre y Dirichlet independientemente) y $\ell = 7$ (Lamé). Así que basta suponer que ℓ es un primo mayor o igual que 11. Así que para demostrar el último teorema de Fermat basta demostrar:

Si $\ell \geq 11$ es primo entonces toda solución entera de $x^\ell + y^\ell + z^\ell = 0$ cumple $xyz = 0$.

A continuación voy a explicar como es la demostración maravillosa que encontraron los matemáticos modernos (con algunas simplificaciones). La demostración es por contradicción:

Sea $\ell \geq 11$ primo y supongamos que existen enteros no nulos a, b, c tales que $a^\ell + b^\ell + c^\ell = 0$. Además, podemos suponer que a, b, c son coprimos, b es par y que $a \equiv 3 \mod 4$ (o sea, 4 divide $a-3$). Esto se puede obtener reordenando a, b, c y simplificando la ecuación por un factor adecuado.

Queremos obtener una contradicción, para poder concluir que dicha solución (a, b, c) no existe.

Idea de partida (G. Frey, 1984). Consideremos la ecuación $E : y^2 = x(x - a^\ell)(x + b^\ell)$. Las raíces del polinomio de la derecha son $0, a^\ell, -b^\ell$ y son distintas porque a^ℓ, b^ℓ, $c^\ell = -a^\ell - b^\ell$ son no nulos. Por lo tanto E es una curva elíptica. El discriminante de esta curva elíptica es

$$\Delta_E = 16((0 - a^\ell)(0 + b^\ell)(a^\ell + b^\ell))^2 = 16(abc)^{2\ell}$$

(¡este es el único lugar donde se usa $a^\ell + b^\ell + c^\ell = 0$ en toda la demostración!) y se puede calcular que el conductor es el producto de los primos que dividen abc, es decir

$$N_E = \prod_{p|abc} p.$$

O sea, el conductor de E es un producto de primos distintos (E es semi-estable) y el discriminante es casi una potencia 2ℓ-ésima. Frey pensó que esta propiedad es muy rara y debería ser incompatible con la modularidad.

[4]Se entiende que hablamos de enteros positivos.

Por el tiempo en que Frey propuso esto, no se sabía que todas las curvas elípticas fueran modulares, y no todos los matemáticos creían que eso fuera posible.

Conjeturas de Serre (J-P. Serre, 1973). En 1973 el genial matemático J-P. Serre propuso una serie de conjeturas en la teoría de las *representaciones de Galois* y formas modulares. Esto fue sin relación al último teorema de Fermat. Cuando más de 10 años después Frey propuso su estrategia para Fermat, Serre observó que si sus conjeturas fueran ciertas entonces la curva elíptica construida por Frey no puede ser modular ¡exactamente como sospechaba Frey! La parte de las conjeturas de Serre necesaria en este contexto era conocida como *conjetura* ϵ. La gente solía decir que para demostrar Fermat había que demostrar *Taniyama-Shimura*+ϵ.

Congruencias entre formas modulares (K. Ribet, 1990): En 1990 Ribet demostró la conjetura ϵ. Más precisamente Ribet demostró (en particular):

Teorema 9.1 (Ribet)**.** *Sea E una curva elíptica semi-estable con discriminante minimal $\Delta_E^{\min} = p_1^{b_1}...p_r^{b_r}$ (factorizado) y conductor N_E. Supongamos que E es modular y que la forma modular asociada es*

$$f = a_E(1)q + a_E(2)q^2 + a_E(3)q^3 + \cdots \in S_2(\Gamma_0(N)).$$

Sea $\ell > 10$ un número primo. Definimos Q como el producto de los p_i tales que ℓ divide b_i y además p_i aparece con exponente 1 en N_E. Entonces hay una forma cuspidal

$$g = c_1 q + c_2 q^2 + c_3 q^3 + \cdots \in S_2(\Gamma_0(N_E/Q))$$

con coeficientes de Fourier enteros tal que para cada n tenemos $a_E(n) \equiv c_n \mod \ell$.

En pocas palabras el contenido del teorema es lo siguiente: *bajo ciertas condiciones, la forma modular asociada a una curva elíptica es congruente módulo cierto primo a otra forma modular pero de* ***nivel más bajo*** *(el nivel baja de N_E a N_E/Q).*

Nota: Cuando Ribet demostró este teorema no se sabía que todas las curvas elípticas fueran modulares, por eso ser modular aparece como condición.

No hemos explicado que cosa es el *discriminante minimal* pero podemos decir lo siguiente: el discriminante minimal $\Delta_E^{\min}$ es "casi" el discriminante Δ_E que calculamos antes, pero es un poco más chico. En nuestro caso es

$$\Delta_E^{\min} = 2^{-8}(abc)^{2\ell}.$$

Apliquemos el teorema de Ribet a nuestro caso. Recordemos que la curva de Frey es semi-estable. Con la información dada y suponiendo que E es modular (para poder aplicar el teorema) se calcula fácilmente $Q = N_E/2$ por lo tanto $g \in S_2(\Gamma_0(2))$ pero este espacio es cero (ver sección 4), así que $g = 0$. Por lo tanto $0 = c_n \equiv a_E(n) \mod \ell$ para todo n. Con $n = 1$ tenemos $1 = a_E(1) \equiv 0 \mod \ell$, una contradicción. Por lo tanto la curva de Frey no puede ser modular, como proponía Frey.

La conjetura de Taniyama-Shimura (Taylor, Wiles): Como ya explicamos en la sección anterior, Wiles y Taylor demostraron en 1995 que todas las curvas elípticas semi-estables son modulares. La curva E construida por Frey es semi-estable así que esta es la última pieza del puzzle. En resumen:

- Supongamos que existen enteros $a, b, c \neq 0$ tal que $a^\ell + b^\ell + c^\ell = 0$.
- *Frey*: Entonces la curva elíptica $y^2 = x(x - a^\ell)(x + b^\ell)$ sería muy rara, parece no ser modular.
- *Serre*: Más precisamente, si mis conjeturas sobre formas modulares y representaciones de Galois fueran ciertas entonces esa curva no puede ser modular.
- *Ribet*: ¡Parte de las conjeturas de Serre son verdad! al menos suficiente para estar seguros que la curva elíptica E (si existiera) definitivamente *no es modular*.
- *Wiles-Taylor*: La conjetura de Taniyama-Shimura es cierta en el caso semi-estable, o sea ¡todas las curvas elípticas semi-estables son modulares! En particular E *es modular*. Contradicción.
- Por lo tanto no existen enteros $a, b, c \neq 0$ tales que $a^\ell + b^\ell + c^\ell = 0$. El último teorema de Fermat está demostrado.

Los trabajos de Ribet [7], Wiles [11] y Taylor y Wiles [10] ocupan más de 170 páginas en total. Puede ser que nunca sepamos que tenía en mente Fermat, pero al menos la demostración moderna tampoco cabe en el margen de una página.

10. ¿Por dónde seguir?

A continuación damos algunas indicaciones para quienes estén interesados en aprender más.

Una introducción a la teoría básica de formas modulares puede encontrarse en [8] (en el caso de $\mathrm{SL}_2(\mathbb{Z})$) y en [4] (en el caso de otros grupos de matrices). El libro [4] tiene ventajas adicionales: tiene una buena introducción a la teoría de curvas elípticas, después revisa las formas modulares de peso entero, y concluye con una introducción a la teoría de formas modulares de peso medio-entero (como la función $\theta(z)$ que tiene peso $1/2$). Para profundizar en el tema de curvas elípticas el libro [9] es la referencia estándar.

El reciente libro [5] es de una lectura liviana y entretenida, introduciendo al lector la relación entre formas modulares y curvas elípticas en el contexto de la conjetura de Taniyama-Shimura y la conjetura de Birch y Swinnerton-Dyer. Destaca por tener ejemplos bien concretos y ayudar al lector a hacer verificaciones por computador.

Las formas modulares también se relacionan con las representaciones de Galois. En el último capítulo de [1] hay una introducción a este tema, que es el primer paso de la demostración de la conjetura de Taniyama-Shimura.

El libro [6] contiene una excelente exposición de varias ideas del área, con énfasis en la aritmética de los coeficientes de Fourier de formas modulares. Está escrito a un nivel accesible y el lector encontrará numerosos problemas abiertos que le darán una idea de cuales son las direcciones actuales en que se desarrolla la teoría.

Se podría seguir hablando interminablemente sobre formas modulares (como diria Lang: esto no es una amenaza), pero en algún momento hay que detenerse.

References

[1] F. Diamond, J. Shurman, *A first course in modular forms*. Graduate Texts in Mathematics, 228. Springer-Verlag, New York, 2005. xvi+436 pp. ISBN: 0-387-23229-X.

[2] R. Hartshorne, *Algebraic geometry*. Graduate Texts in Mathematics, No. 52. Springer-Verlag, New York-Heidelberg, 1977. xvi+496 pp. ISBN: 0-387-90244-9.

[3] M. Kisin, *What is... a Galois representation?* Notices Amer. Math. Soc. 54 (2007), no. 6, 718-719.

[4] N. Koblitz, *Introduction to elliptic curves and modular forms*. Graduate Texts in Mathematics, 97. Springer-Verlag, New York, 1984. viii+248 pp. ISBN: 0-387-96029-5.

[5] A. Lozano, *Elliptic curves, modular forms, and their L-functions*. Student Mathematical Library, 58. IAS/Park City Mathematical Subseries. American Mathematical Society, Providence, RI; Institute for Advanced Study (IAS), Princeton, NJ, 2011. xiv+195 pp. ISBN: 978-0-8218-5242-2.

[6] K. Ono, *The web of modularity: arithmetic of the coefficients of modular forms and q-series*. CBMS Regional Conference Series in Mathematics, 102. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2004. viii+216 pp. ISBN: 0-8218-3368-5.

[7] K. Ribet, *On modular representations of* $\mathrm{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q})$ *arising from modular forms*. Invent. Math. 100 (1990) 431-476.

[8] J-P. Serre, *A course in arithmetic*. Translated from the French. Graduate Texts in Mathematics, No. 7. Springer-Verlag, New York-Heidelberg, 1973. viii+115 pp.

[9] J. Silverman, *The arithmetic of elliptic curves*. Graduate Texts in Mathematics, 106. Springer-Verlag, New York, 1986. xii+400 pp. ISBN: 0-387-96203-4.

[10] R. Taylor, A. Wiles, *Ring-theoretic properties of certain Hecke algebras*. Ann. of Math. (2) 141 (1995), no. 3, 553-572.

[11] A. Wiles, *Modular elliptic curves and Fermat's last theorem*. Ann. of Math. (2) 141 (1995), no. 3, 443-551.

Department of Mathematics and Statistics, Queen's University
Jeffery Hall, University ave.
Kingston, ON Canada, K7L 3N6
E-mail address: hpasten@gmail.com