Representación de enteros como suma de dos cuadrados

Capítulo 4

¿Qué enteros se pueden escribir como suma de dos cuadrados?

Esta pregunta es tan antigua como la teoría de números, y su solución es un clásico del área. La parte "difícil" de la respuesta es comprobar que todo número primo de la forma $4m+1$ es suma de dos cuadrados. G. H. Hardy escribió que este *teorema de los dos cuadrados* de Fermat "es considerado, con toda justicia, como uno de los más bellos de la aritmética." Sin embargo, una de nuestras demostraciones de El Libro es bastante reciente.

Comencemos con algunos preliminares. En primer lugar, hemos de distinguir el primo $p=2$ de los primos de la forma $p=4m+1$ y de los primos de la forma $p=4m+3$. Todo número primo es de uno de estos tres tipos. En este punto, se puede observar (siguiendo la idea de Euclides) que hay infinitos primos de la forma $4m+3$. En efecto, si hubiera sólo un número finito, podríamos tomar el mayor de ellos y llamarlo p_k. Haciendo

$$N_k \;:=\; 2^2 \cdot 3 \cdot 5 \cdots p_k - 1$$

(donde $p_1=2$, $p_2=3$, $p_3=5$, ... denota la secuencia de todos los primos), observamos que N_k es congruente con $3 \pmod 4$ y, por tanto, debe tener un factor primo de la forma $4m+3$. Como dicho factor primo es mayor que p_k, hemos obtenido una contradicción. Al final de este capítulo demostraremos que también hay infinitos primos de la forma $p=4m+1$.

Nuestro primer lema es un caso particular de la famosa "ley de reciprocidad": caracteriza los primos para los cuales -1 es un cuadrado en el cuerpo $\mathbb{Z}_p$ (del que se presenta un resumen en el recuadro de la página siguiente).

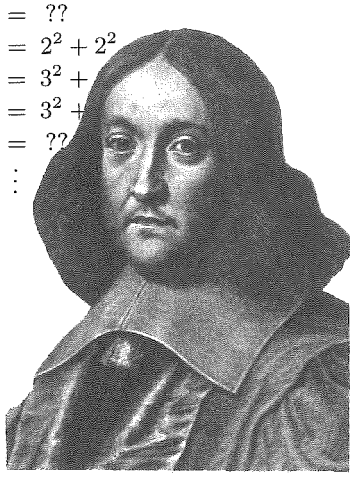
Lema 1. *La ecuación* $s^2 \equiv -1 \pmod p$ *tiene exactamente dos soluciones* $s \in \{1,2,\ldots,p-1\}$ *para los números primos* $p=4m+1$*, una solución para* $p=2$ *y ninguna solución para los primos de la forma* $p=4m+3$.

■ **Demostración.** Para $p=2$ tomamos $s=1$. Si p es impar, consideramos el conjunto $\{1,2,\ldots,p-1\}$ y construimos la relación de equivalencia generada al identificar cada elemento con sus inversos aditivo y multiplicativo en $\mathbb{Z}_p$. Por tanto, la clase de equivalencia "genérica" tiene cuatro elementos

$$\{x,-x,\overline{x},-\overline{x}\},$$

ya que este conjunto contiene los inversos de todos sus elementos. No obstante, hay clases de equivalencia con menos elementos si algunos de los miembros son iguales:

$$\begin{aligned}
1 &= 1^2+0^2\\
2 &= 1^2+1^2\\
3 &= ??\\
4 &= 2^2+0^2\\
5 &= 2^2+1^2\\
6 &= ??\\
7 &= ??\\
8 &= 2^2+2^2\\
9 &= 3^2+\\
10 &= 3^2+\\
11 &= ??\\
&\vdots
\end{aligned}$$



Pierre de Fermat

- $x \equiv -x$ es imposible si p es impar.
- $x \equiv \overline{x}$ es equivalente a $x^2 \equiv 1$. Esta ecuación tiene dos soluciones, $x = 1$ y $x = p-1$, lo que da lugar a la clase de equivalencia $\{1, p-1\}$, con dos elementos.
- $x \equiv -\overline{x}$ es equivalente a $x^2 \equiv -1$. Esta ecuación puede no tener solución o tener dos soluciones distintas, $x_0, p - x_0$. En este último caso, la clase de equivalencia es $\{x_0, p - x_0\}$.

Para $p = 11$ las clases de equivalencia son $\{1, 10\}, \{2, 9, 6, 5\}, \{3, 8, 4, 7\}$; para $p = 13$ son $\{1, 12\}, \{2, 11, 7, 6\}, \{3, 10, 9, 4\}, \{5, 8\}$: el par $\{5, 8\}$ proporciona las dos soluciones de $s^2 \equiv -1 \bmod 13$.

El conjunto $\{1, 2, \ldots, p-1\}$ tiene $p-1$ elementos y lo hemos dividido en 4-uplas (las clases de equivalencia de cardinal 4), más uno o dos pares (las clases de cardinal 2). Para $p - 1 = 4m + 2$ existe sólo el par $\{1, p-1\}$ y el resto son 4-uplas y, por tanto, $s^2 \equiv -1 \,(\mathrm{mod}\, p)$ no tiene solución. Para $p - 1 = 4m$ debe haber un segundo par que contiene las dos soluciones de la ecuación $s^2 \equiv -1$ que estamos buscando. □

Cuerpos primos

Si p es primo, el conjunto $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$ con la suma y multiplicación definidas "módulo p" es un cuerpo finito. Necesitaremos las siguientes propiedades:

- Para $x \in \mathbb{Z}_p$, $x \neq 0$, el inverso aditivo (que denotaremos $-x$) es $p - x \in \{1, 2, \ldots, p-1\}$. Si $p > 2$, entonces x y $-x$ son elementos distintos de $\mathbb{Z}_p$.
- Cada $x \in \mathbb{Z}_p \backslash \{0\}$ tiene un único inverso multiplicativo $\overline{x} \in \mathbb{Z}_p \backslash \{0\}$ tal que $x\overline{x} \equiv 1 \,(\mathrm{mod}\, p)$.
  De la definición de número primo se deduce que la aplicación $\mathbb{Z}_p \to \mathbb{Z}_p$, $z \mapsto xz$ es inyectiva si $x \neq 0$. Por tanto, en el conjunto finito $\mathbb{Z}_p \backslash \{0\}$ debe ser también sobreyectiva, y para cada x debe existir un único $\overline{x} \neq 0$ tal que $x\overline{x} \equiv 1 \,(\mathrm{mod}\, p)$.
- Los cuadrados $0^2, 1^2, 2^2, \ldots, h^2$ son elementos de $\mathbb{Z}_p$ distintos para $h = \lfloor \frac{p}{2} \rfloor$.
  Esto es así porque $x^2 \equiv y^2$ (o $(x+y)(x-y) \equiv 0$) implica que $x \equiv y$ o que $x \equiv -y$. Los $1 + \lfloor \frac{p}{2} \rfloor$ elementos $0^2, 1^2, \ldots, h^2$ se denominan los *cuadrados* de $\mathbb{Z}_p$.

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Suma y multiplicación en $\mathbb{Z}_5$

Mencionemos aquí que para *todos* los números primos existen soluciones de $x^2 + y^2 \equiv -1 \,(\mathrm{mod}\, p)$. De hecho, hay $\lfloor \frac{p}{2} \rfloor + 1$ cuadrados distintos x^2 en $\mathbb{Z}_p$, en tanto que hay $\lfloor \frac{p}{2} \rfloor + 1$ elementos distintos de la forma $-(1 + y^2)$. Como $\mathbb{Z}_p$ tiene p elementos, estos dos conjuntos son demasiado grandes para ser disjuntos, y deben existir x e y tales que $x^2 \equiv -(1 + y^2) \,(\mathrm{mod}\, p)$.

Lema 2. *Ningún número $n = 4m + 3$ es suma de dos cuadrados.*

■ **Demostración.** El cuadrado de un número par es $(2k)^2 = 4k^2 \equiv 0 \,(\mathrm{mod}\, 4)$ y el de un número impar $(2k+1)^2 = 4(k^2 + k) + 1 \equiv 1 \,(\mathrm{mod}\, 4)$. Por tanto, la suma de dos cuadrados es congruente con $0, 1$ ó $2 \,(\mathrm{mod}\, 4)$. □

Ahora que sabemos que los primos de la forma $p = 4m + 3$ son "malos," nos ocuparemos de las "buenas" propiedades de los primos de la forma $p = 4m + 1$. Presentamos el resultado clave previo al teorema principal.

Proposición. *Todo número primo de la forma $p = 4m + 1$ es suma de dos cuadrados, es decir, se puede escribir como $p = x^2 + y^2$ para ciertos números naturales $x, y \in \mathbb{N}$.*

Presentaremos dos demostraciones de este resultado, ambas elegantes y sorprendentes. En la primera se combinan hábilmente el "principio del palomar" (que ya hemos empleado "de pasada" antes del Lema 2; en el capítulo 22 se pueden encontrar más ejemplos), y los argumentos "modulo p". La idea se debe al matemático noruego, especialista en teoría de números, Axel Thue.

■ **Demostración.** Consideremos los pares (x', y') de enteros que verifica que $0 \le x', y' \le \sqrt{p}$, es decir, $x', y' \in \{0, 1, \ldots, \lfloor\sqrt{p}\rfloor\}$. El número total de pares de este tipo es $(\lfloor\sqrt{p}\rfloor + 1)^2$. Utilizando la estimación $\lfloor x\rfloor + 1 > x$ para $x = \sqrt{p}$, vemos que hay más de p tales pares. Por tanto, para cualquier $s \in \mathbb{Z}$, es imposible que todos los valores $x' - sy'$ generados por los pares (x', y') sean distintos módulo p, de donde se deduce que para cualquier s hay dos pares distintos

$$(x', y'),\ (x'', y'') \ \in\ \{0, 1, \ldots, \lfloor\sqrt{p}\rfloor\}^2$$

tales que

$$x' - sy' \ \equiv\ x'' - sy'' \pmod{p}.$$

Restando se tiene que $x' - x'' \equiv s(y' - y'') \pmod{p}$. Si definimos

$$x := |x' - x''|, \quad y := |y' - y''|,$$

se deduce que

$$(x, y) \in \{0, 1, \ldots, \lfloor\sqrt{p}\rfloor\}^2 \quad \text{con} \quad x \equiv \pm sy \pmod{p}.$$

También sabemos que x e y no pueden ser ambos cero, ya que los pares (x', y') y (x'', y'') son distintos.

Sea s una solución de $s^2 \equiv -1 \pmod{p}$, cuya existencia está asegurada por el Lema 1. Como $x^2 \equiv s^2y^2 \equiv -y^2 \pmod{p}$ hemos construido

$$(x, y) \in \mathbb{Z}^2 \quad \text{tal que} \quad 0 < x^2 + y^2 < 2p \quad \text{y} \quad x^2 + y^2 \equiv 0 \pmod{p}.$$

Como p es el único número entre 0 y $2p$ que es divisible por p, se deduce que $x^2 + y^2 = p$, y hemos terminado. □

Para $p = 13$, $\lfloor\sqrt{p}\rfloor = 3$ consideramos $x', y' \in \{0, 1, 2, 3\}$. Para $s = 5$, la suma $x' - sy' \pmod{13}$ toma los siguientes valores:

x' \ y'	0	1	2	3
0	0	8	3	11
1	1	9	4	12
2	2	10	5	0
3	3	11	6	1

La segunda demostración de la proposición — también una demostración de El Libro — fue encontrada por Roger Heath-Brown en 1971 y publicada tan solo en 1984. (Don Zagier dio una versión "resumida en una frase".) Es tan elemental que no necesitaremos utilizar el Lema 1.

El argumento de Heath-Brown utiliza tres involuciones lineales: una que es bastante obvia, otra oculta y una última, trivial, que pone la guinda al pastel. La segunda involución, sorprendente, corresponde a una estructura oculta en el conjunto de soluciones enteras de la ecuación $4xy + z^2 = p$.

Demostración. Estudiaremos el conjunto

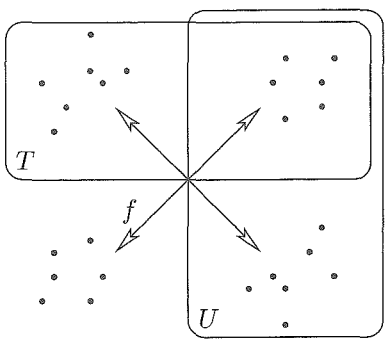
$$S := \{(x,y,z) \in \mathbb{Z}^3 : 4xy + z^2 = p, \quad x > 0, \quad y > 0\}.$$

Este conjunto es finito. De hecho, como $x \geq 1$ e $y \geq 1$ se deduce que $y \leq \frac{p}{4}$ y $x \leq \frac{p}{4}$. Por tanto, hay un número finito de posibles valores de x e y y, para cada uno de ellos, hay como mucho dos valores de z.

1. La primera involución lineal viene dada por

$$f : S \longrightarrow S, \quad (x,y,z) \longmapsto (y,x,-z),$$

es decir, "se permutan x e y, y se cambia el signo a z." Es evidente que esta transformación envía S sobre sí mismo y que es una *involución*: aplicada dos veces, resulta la identidad. Obsérvese que f no tiene puntos fijos, ya que $z = 0$ implicaría que $p = 4xy$. Además, f envía los puntos de



$$T := \{(x,y,z) \in S : z > 0\}$$

a los puntos de $S \backslash T$, que satisfacen $z < 0$. Como f cambia los signos de $x - y$ y de z, manda los puntos de

$$U := \{(x,y,z) \in S : (x-y) + z > 0\}$$

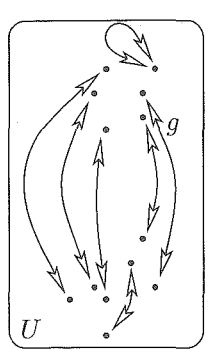
a los puntos de $S \backslash U$. Obsérvese que no pueden existir puntos de S tales que $(x-y)+z = 0$ ya que, en ese caso, se tendría $p = 4xy+z^2 = 4xy+(x-y)^2 = (x+y)^2$.

¿Qué se obtiene del estudio de f? La principal observación es que, como f envía los conjuntos T y U a sus complementarios, también intercambia los puntos de $T \backslash U$ con los de $U \backslash T$. Por tanto, hay el mismo número de puntos en U que no están en T que puntos en T que no están en U — *por tanto, T y U tienen el mismo cardinal.*

2. La segunda transformación que estudiamos es una involución en U:

$$g : U \longrightarrow U, \quad (x,y,z) \longmapsto (x-y+z, y, 2y-z).$$

Veamos, en primer lugar, que g está bien definida: si $(x,y,z) \in U$, entonces $x - y + z > 0, y > 0$ y $4(x-y+z)y + (2y-z)^2 = 4xy + z^2$, por lo que $g(x,y,z) \in S$. Como $(x-y+z) - y + (2y-z) = x > 0$ se deduce que $g(x,y,z) \in U$.



Además, g es una involución: $g(x,y,z) = (x-y+z, y, 2y-z)$ se transforma en $((x-y+z) - y + (2y-z), y, 2y-(2y-z)) = (x,y,z)$.

Finalmente, g tiene exactamente un punto fijo:

$$(x,y,z) = g(x,y,z) = (x-y+z, y, 2y-z)$$

se verifica si y sólo si $y = z$ y, en tal caso, $p = 4xy + y^2 = (4x+y)y$, cuya única solución es $y = 1 = z$ y $x = \frac{p-1}{4}$.

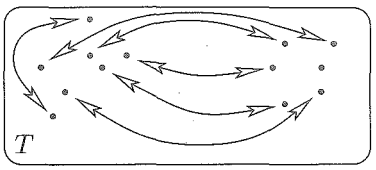
Pero si g es una involución en U que tiene exactamente un punto fijo, entonces *el cardinal de U es impar.*

3. La tercera involución que estudiamos es trivial y consiste en intercambiar x e y:

$$h : T \longrightarrow T, \quad (x, y, z) \longmapsto (y, x, z).$$

Es evidente que esta aplicación está bien definida y que es una involución. Ahora combinamos los hechos deducidos de las otras dos involuciones: el cardinal de T es igual al cardinal de U, que es impar. Pero si h es una involución en un conjunto finito de cardinal impar, entonces *tiene algún punto fijo*: existe un punto $(x, y, z) \in T$ tal que $x = y$, es decir, una solución de

$$p = 4x^2 + z^2 = (2x)^2 + z^2. \qquad \square$$



En un conjunto finito de cardinal impar toda involución tiene al menos un punto fijo.

Obsérvese que esta demostración dice más — el número de representaciones de p de la forma $p = x^2 + (2y)^2$ es *impar* para todos los números primos de la forma $p = 4m + 1$. (De hecho, la representación es única, véase [3].) Obsérvese también que ninguna de las demostraciones es constructiva: ¡intente encontrar x e y para un número primo de diez dígitos! En [1] y [7] se estudian formas eficientes de encontrar tales representaciones como suma de dos cuadrados.

El siguiente teorema responde de forma definitiva a la pregunta con la que comenzó este capítulo.

Teorema. *Un número natural n se puede representar como suma de dos cuadrados si y sólo si todos los factores primos de la forma $p = 4m + 3$ aparecen con exponente par en la factorización de n.*

■ **Demostración.** Diremos que un número n es *representable* si es la suma de dos cuadrados, es decir, si $n = x^2 + y^2$ para ciertos $x, y \in \mathbb{N}_0$. El teorema es una consecuencia de los cinco hechos siguientes.

(1) $1 = 1^2 + 0^2$ y $2 = 1^2 + 1^2$ son representables. Todo número primo de la forma $p = 4m + 1$ es representable.

(2) El producto de dos números representables $n_1 = x_1^2 + y_1^2$ y $n_2 = x_2^2 + y_2^2$ es representable: $n_1 n_2 = (x_1 x_2 + y_1 y_2)^2 + (x_1 y_2 - x_2 y_1)^2$.

(3) Si n es representable, $n = x^2 + y^2$, entonces nz^2 también es representable, ya que $nz^2 = (xz)^2 + (yz)^2$.

Los hechos (1), (2) y (3) proporcionan la parte "si" del teorema.

(4) Si $p = 4m + 3$ es un primo divisor de un número representable $n = x^2 + y^2$, entonces p es divisor de x y de y, por lo que p^2 es divisor de n. De hecho, si $x \not\equiv 0 \pmod p$, existiría $\overline{x}$ tal que $x\overline{x} \equiv 1 \pmod p$, y multiplicando la ecuación $x^2 + y^2 \equiv 0$ por $\overline{x}^2$ obtendríamos la relación $1 + y^2\overline{x}^2 = 1 + (\overline{x}y)^2 \equiv 0 \pmod p$, lo que es imposible según el Lema 1 ya que $p = 4m + 3$.

(5) Si n es representable y $p = 4m + 3$ es un divisor de n, entonces p^2 es divisor de n y n/p^2 es representable. Esto se deduce de (4) y completa la demostración. $\square$

Como corolario, veamos que existen infinitos números primos de la forma $p = 4m + 1$. Para ello, supongamos que hay un número finito y denotemos por p_k al mayor de ellos. Consideremos

$$M_k = (3 \cdot 5 \cdot 7 \cdots p_k)^2 + 2^2,$$

que es congruente con $1 \pmod 4$. Todos sus factores primos son mayores que p_k y, según el hecho (4) de la demostración precedente, no tiene factores primos de la forma $4m + 3$. Por tanto, M_k tiene un factor primo de la forma $4m + 1$ que es mayor que p_k.

Concluimos con dos observaciones:

- Si a y b son dos números naturales que son primos entre sí, entonces hay infinitos números primos de la forma $am+b$ ($m \in \mathbb{N}$) — este es un famoso (y difícil) teorema de Dirichlet. De forma más precisa, se puede demostrar que la cantidad de números primos $p \leq x$ de la forma $p = am + b$ viene dada de forma muy aproximada si x es grande por la función $\frac{1}{\varphi(a)} \frac{x}{\log x}$, donde $\varphi(a)$ denota el número de primos relativos con a que son menores que a. (Esto es una mejora sustancial del teorema de los números primos, que tratamos en la página 10.)
- Esto quiere decir que los números primos para a fijo y b variable aparecen esencialmente en igual cantidad. Sin embargo, por ejemplo para $a = 4$, se puede observar una sutil, pero perceptible y persistente tendencia a "más" primos de la forma $4m + 3$: si observamos un valor de x grande al azar, es probable que haya más primos $p \leq x$ de la forma $p = 4m + 3$ que de la forma $p = 4m+1$. Este efecto se conoce como "desviación de Chebyshev"; véase Riesel [4] y Rubinstein y Sarnak [5].

Referencias

[1] F. W. CLARKE, W. N. EVERITT, L. L. LITTLEJOHN & S. J. R. VORSTER: *H. J. S. Smith and the Fermat Two Squares Theorem,* Amer. Math. Monthly **106** (1999), 652-665.

[2] D. R. HEATH-BROWN: *Fermat's two squares theorem,* Invariant (1984), 2-5.

[3] I. NIVEN & H. S. ZUCKERMAN: *An Introduction to the Theory of Numbers,* Fifth edition, Wiley, New York 1972.

[4] H. RIESEL: *Prime Numbers and Computer Methods for Factorization,* Second edition, Progress in Mathematics **126**, Birkhäuser, Boston MA 1994.

[5] M. RUBINSTEIN & P. SARNAK: *Chebyshev's bias,* Experimental Mathematics **3** (1994), 173-197.

[6] A. THUE: *Et par antydninger til en taltheoretisk metode,* Kra. Vidensk. Selsk. Forh. **7** (1902), 57-75.

[7] S. WAGON: *Editor's corner: The Euclidean algorithm strikes again,* Amer. Math. Monthly **97** (1990), 125-129.

[8] D. ZAGIER: *A one-sentence proof that every prime $p \equiv 1 \pmod 4$ is a sum of two squares,* Amer. Math. Monthly **97** (1990), 144.