Naturales sumas de dos cuadrados

Adolfo Agirre – José Luis Nieva*

1. Introducción

Este artículo se propone analizar el problema de la representación de números naturales como suma de dos cuadrados de números enteros. Sea pues $S = \{0, 1, 4, 9, 25,...\}$ el conjunto de los cuadrados de enteros: $0 = 0^2$, $1 = (+1)^2 = (-1)^2$, etc.. Si se efectúan las sumas de pares de elementos de S (es decir, si se forma el conjunto S + S), se obtienen los números: 0, 1, 2, 4, 5, 8, 9, 10, 13, etc..

Se observa que existen números (3, 6, 7, 11, 12,.....) que no pertenecen a S + S, es decir que no se pueden representar como suma de dos cuadrados.

El problema que se plantea es, por lo tanto, el siguiente:

"Caracterizar los números que pueden representarse como suma de dos cuadrados de enteros".

En el apartado siguiente se verá el teorema fundamental que resuelve este problema, dejando para el último apartado la demostración de un paso clave.

Un recurso básico para esta cuestión es que el conjunto que ha sido llamado S + S es cerrado para el producto.

En efecto, si $m, n \in S + S$, entonces existen enteros a, b, c, d tales que:

$$m = a^2 + b^2 \quad ; \quad n = c^2 + d^2.$$

Sean los números complejos α y β dados por:

$$\alpha = a + ib \quad ; \quad \beta = c + di$$

y sea $N(\alpha) = a^2 + b^2$, la norma de α (igual cuadrado del módulo). Si se aplica una propiedad muy conocida de los complejos, a saber: $N(\alpha\beta) = N(\alpha)N(\beta)$, se tiene:

$$m\,n = N(\alpha)N(\beta) = N(\alpha\beta) \quad , \quad \text{y como}$$

$\alpha\beta = (ac - bd) + (ad + bc)i$, resulta : $m\,n = (ac - bd)^2 + (ad + bc)^2 \in S + S$.

2. Teorema de Caracterización

La propiedad de cierre para el producto de S + S, sugiere que se analice previamente el caso de los números primos.

Si p = 2, entonces $p \in S + S$, pues $2 = 1^2 + 1^2$.

Si p es impar y se lo divide por 4, el resto es 1 o 3.

Si p es de la forma 4k + 3, o sea $p \equiv 3 \pmod 4$, es imposible que $p \in S + S$. En efecto, cualquier número x es congruente módulo 4 con 0, 1, 2, o 3 y su cuadrado x^2 es congruente con 0 o 1 (mód 4). Por lo tanto una suma de dos cuadrados $x^2 + y^2$ sólo puede ser congruente con 0, 1 o 2 (mód 4). Luego $p = 4k + 3 \notin S + S$.

Si p es de la forma 4k + 1, o sea $p \equiv 1$ (mód 4), se demuestra que $p \in S + S$. Este hecho fundamental será probado en el apartado siguiente. Será admitido en lo que resta de este apartado.

Pasando al análisis de los números compuestos, una vez factorizados en sus factores primos, podrán escribirse en la forma:

$$n = 2^j p_1^{a_1} \ldots\ldots p_k^{a_k} q_1^{b_1} \ldots\ldots q_r^{b_r}$$

siendo los primos p_k de la forma 4m + 1 y los primos q_r de la forma 4m + 3.

Si el exponente b_r de q_r es par, entonces $q_r^{b_r}$ es un cuadrado perfecto u^2 y pertenece a S + S bajo la forma $u^2 + 0^2$.

De lo anterior resulta que, si todos los b_r son pares, entonces $n \in S + S$. En el resto de este apartado se prueba la validez del recíproco (o de su contrario, que es equivalente).

Primero se analiza un caso particular. Sea q primo de la forma 4m + 3 que divide a n. Se va a probar que n no admite una representación $n = x^2 + y^2$ con x e y coprimos.

Se supone pues que: $n = x^2 + y^2$, $(x , y) = 1$.

Si $q \mid x$, también será $q \mid y$, lo que es imposible por ser x e y coprimos.

Siendo que q no divide a x, q primo, entonces x y q son coprimos y por lo tanto existen enteros u, v tales que:

$$ux + vq = 1 \text{ , de donde } ux \equiv 1 (\text{mód } q)$$

De $n = x^2 + y^2$ y $q \mid n$ se deduce que: $x^2 + y^2 \equiv 0$ (mód q).

Multiplicando por u^2:

$$u^2 x^2 + u^2 y^2 \equiv 0 (\text{mód } q) \quad \text{o sea } 1 + (uy)^2 \equiv 0 (\text{mód } q).$$

Llamando z a uy nos queda $1 + z^2 \equiv 0$ (mód q).

O sea: $-1 \equiv z^2$ (mód q) , lo cual es imposible, pues – 1 no es resto cuadrático de los primos de la forma 4m + 3.

Sea ahora q un factor primo de n que aparece en la factorización de n a un exponente b impar. Se va a mostrar que $n \notin S + S$. Sea $n = x^2 + y^2$, y sea d el m.c.d. de x e y: $d = (x, y)$. Llamando x_0 e y_0 a los cocientes de x e y por d, se tiene:

$$x = dx_0 \quad , \quad y = dy_0 \quad , \quad (x_0 , y_0) = 1 .$$

Sea g la potencia (≥ 0) a la que figura q en d.

De $n = x^2 + y^2 = d^2 (x_0^2 + y_0^2)$, resulta que $d^2 \mid n$; sea n_0 el cociente de n por d^2.

En n_0 el primo q figura elevado al exponente b – 2g, el cual es positivo por ser b impar.

Se tiene pues: $n_0 = x_0^2 + y_0^2$, $q \mid n_0$, $(x_0 , y_0) = 1$ contradiciendo lo que se demostró anteriormente.

Ha quedado así probado el siguiente teorema que caracteriza a los elementos de S + S (ver (HW) o (BPS)):

Teorema: Un número natural n se puede representar como la suma de dos cuadrados de números enteros si y sólo si en su expresión como producto de números primos, los primos de la forma 4k + 3 (si los hubiera) aparecen con exponente par.

Por ejemplo: $1300 = 2^2 . 5^2 . 13$ y $49000 = 2^3 .5^3 .7^2$ son ambos representables como suma de dos cuadrados, en cambio $630 = 2 .3^2 . 5 .7$ no lo es.

3. Primos de la forma 4k+1

Es necesario probar el siguiente teorema:

Teorema: Si p es primo de la forma $4k + 1$, entonces existen enteros x e y tales que $p = x^2 + y^2$.

Comentario previo:
Hay dos demostraciones muy difundidas:

1) Con el método del "Descenso infinito", usado por Euler en la primera prueba publicada y, que habría sido el de Fermat. Se prueba que si un múltiplo mp de p es representable como una suma de dos cuadrados, lo mismo ocurre para un múltiplo menor m'p, con $0 < m' < m$. Como los m son naturales, es imposible un descenso infinito, es decir, luego de un número finito de pasos se llega a $1p = p = x^2 + y^2$. Esto puede verse en (HW) o (D).
2) Con el uso de enteros gaussianos, es decir complejos de componentes enteras (simbolizado por Z[i]). Se puede definir en Z[i] una división con resto similar a la de $\wedge$ y con un argumento de divisibilidad se llega a probar el teorema, mostrando que hay un entero gaussiano $\alpha = x + iy$, cuya norma es p, $p = N(\alpha) = x^2 + y^2$. Puede verse en (H).

Otra demostración interesante, usando el principio de los casilleros, puede verse en (BPS).
Se va a exponer aquí la demostración de Zagier (Z), que sorprende por los métodos que emplea y que a diferencia de las antes mencionadas, no necesita el hecho de –1 sea resto cuadrático de p.

Demostración: Se trabaja con el conjunto A de ternas de números naturales (x,y,z) que verifican la relación: $x^2 + 4yz = p$.

Por ser p primo, no pueden ser cero ni x, ni y, ni z.

La variable x no puede tomar valores pares, pues en tal caso $p - x^2$ sería impar. Puede en cambio tomar cualquier valor impar entre 1 y el mayor impar x_o cuyo cuadrado sea menor que p. Como p y x^2 son congruentes con 1 módulo 4, $p - x^2$ es múltiplo de 4 y el cociente $\frac{p-x^2}{4}$ se puede factorear en la forma yz en un número finito de formas.

Por lo tanto el conjunto (no vacío) A es finito.

Conviene recordar ahora lo que se entiende por involución en un conjunto S. Una función biyectiva f de S sobre S es una involución si coincide con su inversa f^{-1}: $f^{-1} = f$, o lo que es lo mismo, que compuesta consigo misma es la identidad de S: $f \circ f = Id_S$ .Es preferible esta segunda definición pues de ella se deduce que f es biyectiva.

Ejemplos importantes de involuciones los tenemos en la geometría plana donde las simetrías centrales y axiales son involuciones.

La idea de la prueba es demostrar primero que #A (cardinal de A) es impar. Una vez hecho esto, se considera la función $g : A \to A$ definida por $g(x,y,z) = (x,z,y)$ que evidentemente es una involución. La involución g permuta entre sí las ternas con $y \neq z$; estas ternas tienen un cardinal par, por lo tanto deben quedar ternas con $y = z$ y que son puntos fijos de g. Un tal punto fijo de g : (x,y,y) nos da la expresión buscada, pues :

$x^2 + 4y^2 = p$, o sea $p = x^2 + (2y)^2$.

(Nota :Hay una sola terna (x,y,y), pero no se necesita probarlo, ya que sólo interesa la existencia de solución, no la unicidad).

Para mostrar que #A es impar, Zagier usa otra involución f de A que tiene un solo punto fijo.

Como $y - z \geq 2y$ equivale a $0 \geq y + z$ que es imposible, el conjunto A se puede descomponer en tres subconjuntos, según la ubicación de x respecto a estos elementos $y - z$, $2y$:

$B = \{(x,y,z) \in A / x < y - z\}$
$C = \{(x,y,z) \in A / y - z < x < 2y\}$
$D = \{(x,y,z) \in A / 2y < x\}$

Como tanto: 1) $x = y - z$, que implica $(y + z)^2 = p$
2) $x = 2y$, de donde $4y^2 + 4yz = p$, luego $4 \mid p$

son imposibles por ser p primo, B, C y D son disjuntos dos a dos y su unión es A. Salvo para k = 1, p = 5, caso en el cual B y D son vacíos y A = C = {(1,1,1)}, en los demás casos,
$k \geq 3$ (k = 2 no da primo), se tienen : $(1,k,1) \in B$, $(3,1,k-2) \in D$, $(1,1,k) \in C$, y por lo tanto {B,C,D} es una partición de A.
Se define f en A en la siguiente forma:

$$f(x,y,z) = \begin{cases} (x + 2z, z, y - x - z) & \text{si } (x,y,z) \in B \\ (2y - x, y, x - y + z) & \text{si } (x,y,z) \in C \\ (x - 2y, x - y + z, y) & \text{si } (x,y,z) \in D \end{cases}$$

De inmediato se ve que las condiciones que definen B , C y D permiten asegurar que las componentes de f(x,y,z) son naturales no nulos.
Ahora hay que verificar que f aplica B en D, C en C y D en B.
Sea $(x,y,z) \in B$, o sea $x^2 + 4yz = p$, $x < y - z$. Como $f(x,y,z) = (x + 2z, z, y - x - z)$, se comprueba que está en A en forma directa: $(x + 2z)^2 + 4z(y - x - z) = x^2 + 4xz + 4z^2 + 4yz - 4xz - 4z^2 = x^2 + 4yz = p$. Está en D pues: $2z < x + 2z$, o sea $x > 0$.
Sea $(x,y,z) \in C$, o sea $x^2 + 4yz = p$, $y - z < x < 2y$. Igual que antes: $f(x,y,z) = (2y - x, y, x - y + z)$, $(2y - x)^2 + 4y(x - y + z) = 4y^2 - 4xy + x^2 + 4xy - 4y^2 + 4yz = x^2 + 4yz = p$; está en C pues: $y - (x - y + z) < 2y - x$, o sea $z > 0$, $2y - x < 2y$, o sea $x > 0$.
Sea $(x,y,z) \in D$, o sea $x^2 + 4yz = p$, $2y < x$. Igual que antes: $f(x,y,z) = (x - 2y, x - y + z, y)$, $(x - 2y)^2 + 4(x - y + z)y = x^2 - 4xy + 4y^2 + 4xy - 4y^2 + 4yz = x^2 + 4yz = p$; está en B pues: $x - 2y < x - y + z - y$, o sea $z > 0$.
Ahora se comprueba que f es involución :
1°) $(x,y,z) \in B$, entonces: $f(f(x,y,z)) = f(x + 2z, z, y - x - z) =$ (se aplica la expresión correspondiente a D) $= (x + 2z - 2z, x + 2z - z + y - x - z, z) = (x,y,z)$
2°) $(x,y,z) \in C$, entonces $f(f(x,y,z)) = f(2y - x, y, x - y + z) =$ (se aplica la expresión correspondiente a C) $= (2y - 2y + x, y, 2y - x - y + x - y + z) = (x,y,z)$
3°) $(x,y,z) \in D$, entonces $f(f(x,y,z)) = (x - 2y, x - y + z, y) =$ (se aplica la expresión correspondiente a B) $= (x - 2y + 2y, y, x - y + z - x + 2y - y) = (x,y,z)$.
Luego f es biyectiva. Más precisamente, se establece un biyección entre B y D y una biyección de C sobre C.
La terna $(1,1,k) \in C$, pues $1 - k \leq 0 < 1 < 2$.
Su imagen es $f(1,1,k) = (1,1,k)$ o sea $(1,1,k)$ es punto fijo de f.

Sea (x,y,z) $\in$ C punto fijo de f, o sea $2y - x = x$, $y = y$, $x - y + z = z$. De la primera resulta $x = y$, por lo tanto $x^2 + 4xz = p$, o sea $x(x + 4z) = p$, de donde $x = 1$ por ser p primo y como $p = 4k + 1$, resulta $z = k$.
Luego el único punto fijo de f es (1,1,k). Como f es una involución de A con un úni-co punto fijo, entonces #A es impar. Como se dijo antes, esto concluye la demostración.

Bibliografía

(BPS): Becker-Pietrocola-Sánchez: Aritmética – Red Olímpica – 2001 – Cap. 7.

(D): Davenport: The Higher Arithmetic. Hutchinson University Library. Londres 1970. Cap V.

(H): Herstein: Algebra moderna – Edit. Trillas. México.1970. Cap. 3.

(HW): Hardy-Wright: An Introduction to the Theory of Numbers – Oxford University Press. Londres. 5° edición. 1965. Cap. XX.

(Z): Zagier: "A one-sentence proof that every prime $p \equiv 1$ (mód 4) is a sum of two squares" – American Mathematical Monthly (1990) Pág. 144.

*Caseros Norte 422 – 4700 Catamarca.

Departamento de Matemática
Facultad de Ciencias Exactas.
Universidad Nacional de Catamarca.