JORNADAS SOBRE LOS
PROBLEMAS DEL MILENIO
Barcelona, del 1 al 3 de junio de 2011

LA CONJETURA DE BIRCH Y SWINNERTON-DYER

VÍCTOR ROTGER

Introducción

Recuerdo muy bien la primera vez que oí hablar de la conjetura de Birch y Swinnerton-Dyer. Fue unos meses antes de acabar la licenciatura en la Universitat de Barcelona, en junio de 1998.

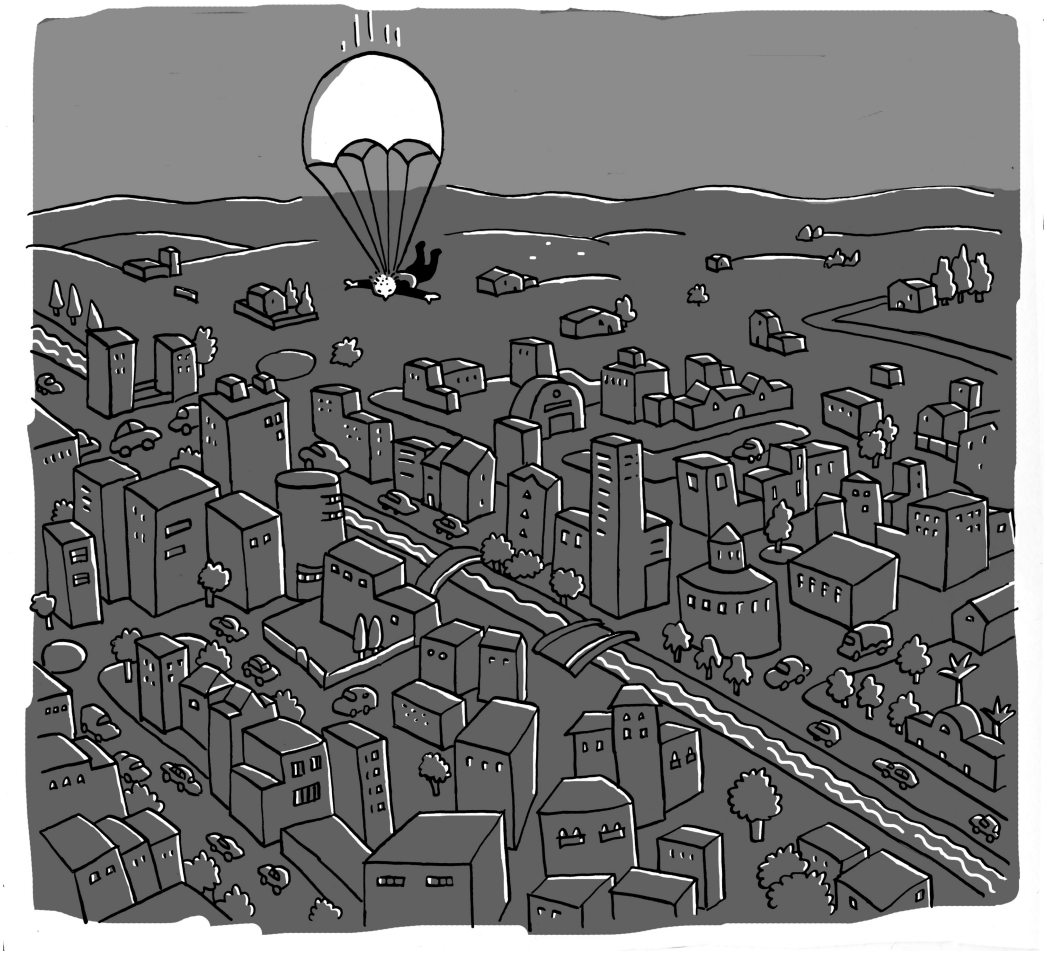
Había llamado a la puerta de la profesora Pilar Bayer para preguntarle si querría ser la directora de mi tesis doctoral. Ella me abrió con una cálida sonrisa y me hizo entrar en su despacho, que tenía repleto de libros y armarios rebosantes de papeles y artículos. Me sugirió que a lo mejor podría introducirme en la teoría de números a través del estudio de unas curvas que se llaman *curvas de Shimura*. Acompañando sus palabras, hizo un gesto inconsciente con el dedo, como si dibujara en el aire una de ellas.

A continuación me explicó que las curvas de Shimura juegan un rol importante en la demostración del último teorema de Fermat que tan recientemente había encontrado Andrew Wiles, y que sin duda serían una pieza clave en cualquier avance que pudiera haber en el futuro sobre la conjetura de Birch y Swinnerton-Dyer, que aún permanecía completamente abierta.

No sé cuál debía ser mi expresión durante los minutos que estuve en el despacho de Pilar Bayer esa primera vez, pero yo me sentía como un paracaidista precipitándome sobre una ciudad en la que nunca antes había estado.

Desde entonces han transcurrido trece años, y la conjetura de Birch y Swinnerton-Dyer permanece intacta. Estas notas son una introducción a esta conjetura y a alguna de las ideas que en los años sesenta llevaron a los dos matemáticos británicos que le dan el nombre a formularla.

Hay muchísima gente que me ha ayudado a adentrarme en los entresijos de la conjetura de Birch y Swinnerton-Dyer. En primer lugar está Pilar Bayer, mi directora de tesis doctoral, y todos mis compañeros del grupo de teoría de números de la Universitat Politècnica de Catalunya, la Universitat de Barcelona y la Universitat Autònoma de Barcelona. Además de ellos, también están los matemáticos con quienes he colaborado durante los últimos años: querría destacar entre ellos a Henri Darmon, con quien cada conversación abre las puertas a una idea nueva y prometedora o construye un puente entre dos ideas sin aparente relación entre sí. Finalmente, estas notas no serían lo que son sin la ayuda de algunas personas muy cercanas a mí. Este manuscrito va dedicado a ellas.



1. Ecuaciones diofánticas

Resolver ecuaciones diofánticas, aunque no las llamemos así, lo llevamos haciendo todos desde que aprendimos a contar. De hecho, seguramente *aprendimos* a contar jugando a resolver ecuaciones diofánticas. Y ese juego se lleva practicando desde hace muchos siglos, sin duda mucho antes que el mismo Diofanto. Sería sólo después de la publicación de su libro *Aritmética*, en el que recopiló, resolvió y planteó muchas de estas ecuaciones, dejando muchas de ellas sin solución, que se acuñó el término *ecuación diofántica*.

Resolver una ecuación de éstas no es otra cosa, por poner un ejemplo, que decidir cuántas magdalenas me voy a comer mientras miro una película si en el

armario sólo me queda una bolsa de diez, es domingo por la tarde con lo que los supermercados están cerrados y mañana lunes por la mañana *necesito* como mínimo cuatro para desayunar. La respuesta es evidentemente 6, diga lo que diga Diofanto. Lo único que aportaría Diofanto y su escuela en este caso es plantear la ecuación

$$x + 4 = 10, \tag{1}$$

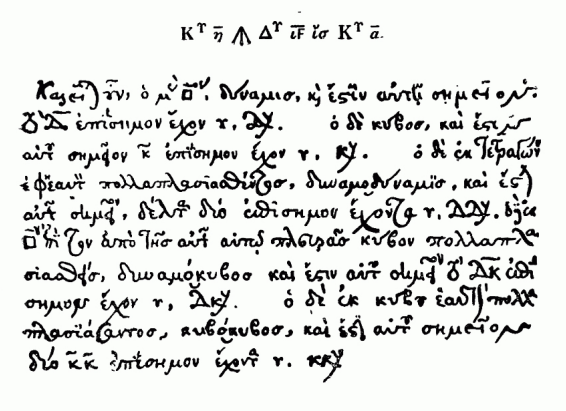
cosa que en realidad no nos importa demasiado mientras nos podamos comer las magdalenas tranquilamente.

Así que una ecuación diofántica es entre otras cosas lo que la gente simplemente llama una *ecuación*. Cómo tiene que ser la ecuación para que se considere *diofántica*?

Por ecuación diofántica muchas veces la gente simplemente piensa en una ecuación sencilla y fácil de resolver, con números sencillos y fáciles de entender. O mejor dicho, como a veces lo que importa en este mundo sólo son las apariencias, una ecuación que *parezca* sencilla y fácil de resolver. Pero todos sabemos que por desgracia no es lo mismo ser que parecer.

La verdad es que en los libros uno encuentra varios conceptos diferentes bajo el paraguas del término *ecuaciones diofánticas*, así que tampoco nos va a preocupar mucho aquí el definir rigurosamente el término: más bien nos contentaremos con algunos ejemplos sencillos y con entender por qué otros pueden ser asombrosamente difíciles.

Eso es lo que indudablemente debía atraer al mismo Diofanto: cómo es posible que dos ecuaciones diferentes, pero muy parecidas, puedan entrañar niveles de dificultad tan distintos. En este capítulo nos dedicaremos a profundizar en esta cuestión. Por el momento, ahí abajo va un párrafo de su libro, en su versión griega original, que tanto inspiró a matemáticos posteriores como Carl F. Gauss o Pierre de Fermat, y sigue inspirando a muchos otros en la actualidad.



La *Arimética* de Diofanto de Alejandría.

1.1. Conjuntos de números. Veamos y comparemos algunas ecuaciones diofánticas, para que se sepa de qué estamos hablando. La ecuación (1) de arriba, por ejemplo, parece fácil y *lo es*. Todos sabemos resolver (1) incluso sin necesidad de plantearla. Así que todos estamos de acuerdo en que (1) es una ecuación diofántica. Y puestos a poner ejemplos con los mismos números, podemos plantear las siguientes ecuaciones:

$$x + 10 = 4, \tag{2}$$

$$4x = 10, \tag{3}$$

$$x^4 = 10. \tag{4}$$

Y aquí uno ya ve que la cosa se puede complicar sin comerlo ni beberlo.

Resolver (2), por ejemplo, sigue siendo fácil: pasas el 10 al otro lado y te queda $x = 4 - 10 = -6$. Así que la solución es $x = -6$. Quizás la única diferencia con la ecuación (1) es que esta vez ya no sé muy bien cómo plantear el tema de las magdalenas: seguro que he resuelto esta ecuación muchas menos veces en casa cuando las buscaba en el armario.

Así que todos estaremos de acuerdo en que la solución de la ecuación (2) sigue siendo igualmente fácil pero menos natural, porque en la calle se oye más el 6 que el -6. Y todos nos acordamos (aunque para unos hace más años de eso que para otros) de cuando decíamos que (2) *no tiene solución* porque los números negativos no existen.

Ahora que ya somos mayores decimos que tanto el 6 como el -6 son números *enteros*, tan bueno el uno como el otro, sólo que el 6 es un número natural mientras que el -6 no. En símbolos, el conjunto de los números naturales se denota

$$\mathbb{N} = \{1, 2, 3, 4, 5...\}$$

y el de los enteros

$$\mathbb{Z} = \{..., -4, -3, -2, -1, 0, 1, 2, 3, 4...\}.$$

En cambio también estaréis de acuerdo en que las ecuaciones (1) y (2) parecen y hasta son igual de naturales: sólo salen los números x, 4, 10 sumándose.

El segundo es un ejemplo sencillo de ecuación cuyos coeficientes son todos naturales pero que no admite soluciones naturales. Es bastante común, pues, encontrarse ecuaciones de apariencia sencilla pero de solución más complicada. Como las personas, así que tampoco nos extraña demasiado.

Vayamos por la siguiente: todos convenimos en decir que la ecuación (3) también es muy fácil: pasamos el 4 dividiendo abajo y nos queda $x = \frac{10}{4}$, que a lo mejor

podemos simplificar y obtener $x = \frac{5}{2}$ o hasta poner en forma decimal si queremos: $x = 2{,}5$. Comoquiera que la expresemos, ésta es la (única) solución.

De nuevo (3) es una ecuación cuyos coeficientes son números naturales pero su solución no es natural, y en este caso ni siquiera entera: $x = \frac{5}{2}$ es lo que todos llamamos un número *racional*.

Esta vez, pues, hemos tenido que ampliar el conjunto de soluciones hasta el de los números racionales para poder encontrarla. Simbólicamente nos referiremos a este conjunto como

$$\mathbb{Q} = \{0, -6, 13, \frac{2}{3}, \frac{-7}{11}, \frac{11}{12}, ...\} = \{\frac{m}{n} \text{ donde } m, n \in \mathbb{Z} \text{ y } n \neq 0\}.$$

Finalmente, la ecuación (4) tiene ya una pinta menos inocua, aunque sigue siendo verdad que sólo hemos utilizado los números x, 4 y 10 y las operaciones de sumar y multiplicar. Dicho en voz alta, se nos pide encontrar un número (un número *¿qué?*) x tal que $x \cdot x \cdot x \cdot x$ sea igual a 10. Vamos, la raíz cuarta de 10.

A bote pronto a uno no se le ocurre ninguno: $x = 1$ no sirve porque $1^4 = 1$ y con $x = 2$ queda $2^4 = 16$, que ya se pasa, así que no parece que vayamos a encontrar la solución entre los números enteros. Si empezamos a probar con los números racionales, uno puede encontrar algunas aproximaciones bastante buenas. Por ejemplo $x = \frac{16}{9}$ da

$$x^4 = 9{,}988721231519585429050449626 58...$$

que no está mal, aunque no es 10. Sea como sea, este es el número racional positivo con denominador menor que 100 que mejor se aproxima a $\sqrt[4]{10}$.

Y si empiezas a probar con denominadores más grandes, hasta 1000 por ponerse un límite razonable, descubrirás (al cabo de *bastante rato*, si tu estrategia es la de probar al tuntún) que el número racional que mejor se aproxima a la solución verdadera es $x = \frac{393}{221}$. Concretamente, da

$$x^4 = 10{,}000025484152351975372748340 8...$$

que está bastante mejor que el anterior.

Para encontrar estas aproximaciones hay maneras mejores que probar uno a uno todos los números racionales que hay con denominador menor que 1000: yo por ejemplo he utilizado el método de las *fracciones continuas*, que no explicaremos aquí pero que es muy curioso y sencillo: investigadlo. Diofanto mismo ya lo conocía.

Lo cosa está en que por mucho que lo intentemos, no hay ningún número racional $x = \frac{m}{n}$ que sea solución de (4). (¿Por qué? Demostradlo.) Lo mejor que sabemos hacer es encontrar más y mejores aproximaciones racionales a la solución.

Y con la misma frescura que uno se inventó los números negativos para que (2) no se quedara sin respuesta, o los racionales para que (3) tuviera solución, uno va y simplemente dice que una solución de (4) *es* $\sqrt[4]{10}$. Por la cara.

Aunque parezca un poco extraño, eso es lo bueno de las matemáticas: uno puede inventarse todo lo que quiera siempre, mientras convenza al resto de que ese nuevo objeto no contradice ninguno de los anteriores.

Hay un cierto grado de ambigüedad en el símbolo $\sqrt[4]{10}$: observad que sólo he dicho que es *una* solución, no *la* solución. Y es que si damos en llamar $\sqrt[4]{10}$ a una de ellas, entonces por ejemplo $-\sqrt[4]{10}$ también lo es, ya que $(-\sqrt[4]{10})^4 = (-1)^4(\sqrt[4]{10})^4 = 10$, y una no es mejor que la otra.

Muchos ya se quedan satisfechos con ese símbolo como una de las soluciones de (4), entre los cuales me incluyo bastante a menudo. A otros les gusta ser más precisos y dicen que una solución de (4) es el *límite* de las aproximaciones racionales positivas

$$\frac{16}{9}, \frac{393}{221}, \frac{7090}{3987}, \frac{227273}{127805}... \longrightarrow \text{solución de (4)} \tag{5}$$

que uno encuentra al permitir denominadores más y más grandes. Ese límite en realidad *no existe* dentro del conjunto de los números racionales, así que lo llamemos como lo llamemos, estamos inventándonos números nuevos para poder calcular las soluciones de (4).

Los números obtenidos de esta manera, como límite de racionales, se llaman *números reales* y su conjunto se denota con el símbolo $\mathbb{R}$.

En el ejemplo que estamos examinando, podemos bautizar con el símbolo $\sqrt[4]{10}$ al número que encontramos como límite de la sucesión (5). Es pues un número *real* y es de hecho la única solución real positiva de la ecuación $x^4 = 10$. Como también es verdad que $\pm\sqrt[4]{10}$ son las dos únicas soluciones reales de (4). Concretamente, $\sqrt[4]{10}$ y $-\sqrt[4]{10}$ son números reales que no son racionales, así que a menudo se dice que $\pm\sqrt[4]{10}$ son *irracionales* y se escribe

$$\sqrt[4]{10}, -\sqrt[4]{10} \in \mathbb{R} \setminus \mathbb{Q}.$$

Puestos a inventar, si uno introduce incluso el conjunto de los números *complejos*

$$\mathbb{C} = \{a + bi,\ a, b \in \mathbb{R}\}$$

donde convenimos que el símbolo i es un *número* nuevo que uno se inventa decretando sin más que $i^2 = -1$, entonces todos estaremos de acuerdo –por decreto– en que los números

$$\{\sqrt[4]{10}, -\sqrt[4]{10}, \sqrt[4]{10}i, -\sqrt[4]{10}i\}$$

son todos ellos soluciones diferentes de (4). Para comprobarlo para $\sqrt[4]{10}i$, por ejemplo, sólo hace falta utilizar las reglas establecidas: $(\sqrt[4]{10}i) = (\sqrt[4]{10})^4 i^4 = 10(-1)^2 = 10$.

El *Teorema fundamental del álgebra* (que refresco con más detalle un poco más abajo) nos garantiza que esas son efectivamente todas las posibles soluciones de la ecuación dentro del conjunto $\mathbb{C}$: hay tantas como el *grado* de (4).

Los diferentes conjuntos de números que han aparecido hasta el momento están sucesivamente incluidos uno dentro del otro, puesto que los introdujimos de manera paulatina al vernos obligados a dar solución a las ecuaciones diofánticas (1), (2), (3) y (4). Simbólicamente lo representamos así:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}. \tag{6}$$

En cualquier caso, en este texto no vamos a profundizar mucho sobre quiénes son los números reales y complejos, y nos quedaremos con los naturales, enteros y racionales, que nos son más próximos y ya plantean suficientes interrogantes de esos que *parecen* fáciles pero que no lo son.

1.2. Clasificación de ecuaciones polinómicas. Convengamos, para centrar un poco la discusión, que una ecuación diofántica es para nosotros una ecuación

$$E : f = 0$$

donde f es un *polinomio* de una o más variables (x, y, z, ...) con *coeficientes* en un conjunto de números dado.

En los ejemplos anteriores, todas las ecuaciones son de *una única* variable x, con coeficientes en $\mathbb{N}$. La ecuación (3), por ejemplo, viene dada por el polinomio $f(x) = 4x - 10$, que tiene *grado* 1. Y el polinomio de la ecuación (4) es $f(x) = x^4 - 10$, que tiene grado 4. Otros ejemplos de ecuaciones diofánticas son

(7)

$E : f = 0$	Variables	Grado	Coeficientes
$x^3 - 11x + 4 = 0$	x	3	$\mathbb{Z}$
$x^7 - \frac{1}{2}x = 0$	x	7	$\mathbb{Q}$
$x + \sqrt{2} = 0$	x	1	$\mathbb{R}$
$x^2 + y^2 + 1 = 0$	x, y	2	$\mathbb{N}$
$x^2 + xy + y^2 = 0$	x, y	2	$\mathbb{N}$
$y^2 - x^3 - x - 1 = 0$	x, y	3	$\mathbb{Z}$
$3x^{13} + 4y^{12} - 5z^{11} = 0$	x, y, z	13	$\mathbb{Z}$

Aclaremos antes de seguir qué entendemos exactamente por el *grado* y el *conjunto de coeficientes* de una ecuación, porque jugarán un rol importante en las siguientes páginas.

El grado de un polinomio formado por un único término como por ejemplo

11	x	y	$3x^2$	$2xy$	$-x^3y$	$2x^2yz^2$

es, respectivamente,

0	1	1	2	2	4	5

Para calcularlo uno sencillamente debe sumar los exponentes de cada una de las variables. Si un polinomio tiene varios términos, su grado es entonces el máximo de ellos. Es por esta razón que en la tabla anterior obtenemos que el polinomio $y^2 - x^3 - x - 1$ tiene grado 3, y $3x^{13} + 4y^{12} - 5z^{11}$ tiene grado 13.

El cuanto al conjunto de coeficientes de la ecuación, con él nos referimos a *cualquier* conjunto que contenga a todos los coeficientes del polinomio que la define. En la tabla anterior hemos consignado el *menor* conjunto de coeficientes de cada una de las ecuaciones de entre los cinco que hemos considerado en (6). Pero no sería ninguna contradicción decir que la ecuación $3x^{13} + 4y^{12} - 5z^{11} = 0$ tiene

coeficientes en $\mathbb{Q}$ o en $\mathbb{R}$, por decir dos posibles conjuntos, puesto que *es cierto* que sus coeficientes $3, 4$ y -5 son números racionales y también reales. En cambio sería absurdo afirmar que la ecuación $x + \sqrt{2} = 0$ tiene sus coeficientes en $\mathbb{Q}$, porque $\sqrt{2}$ es un número irracional.

Una vez hechas estas aclaraciones, para seguir avanzando es importante subrayar el fenómeno con el que hemos topado: una ecuación diofántica puede *no tener soluciones* sobre su mismo conjunto de coeficientes y que tengamos que buscarlas en un conjunto mayor.

Es el caso de la ecuación (4), cuyos coeficientes están en el conjunto $\mathbb{N}$ de los números naturales, pero no tiene ninguna solución en $\mathbb{N}$, $\mathbb{Z}$ o $\mathbb{Q}$. Cuando ampliamos nuestro ámbito de soluciones al conjunto $\mathbb{R}$ de los números reales, encontramos dos de ellas. Y finalmente en el conjunto $\mathbb{C}$ de los números complejos es donde encontramos dos más. En otras palabras, si llamamos

$$E : x^4 - 10 = 0$$

a la ecuación planteada en (4) y para cualquier conjunto C de números escribimos

$$E(C) = \{x \in C \text{ tales que } x^4 - 10 = 0\}, \tag{8}$$

hemos visto que

$$E(\mathbb{N}) = E(\mathbb{Z}) = E(\mathbb{Q}) = \emptyset$$

es el conjunto vacío y que

$$E(\mathbb{R}) = \{\pm\sqrt[4]{10}, -\sqrt[4]{10}\} \subset E(\mathbb{C}) = \{\pm\sqrt[4]{10}, \pm\sqrt[4]{10}i\}.$$

Tales ejemplos se pueden construir a mansalva, con los conjuntos $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ y $\mathbb{C}$ jugando casi cualquiera de los roles de existencia o no-existencia de soluciones, mientras respetemos las cadenas obvias de inclusiones.

De forma paralela a (6), los conjuntos de soluciones de una ecuación diofántica E están sucesivamente incluidos uno dentro del otro:

$$E(\mathbb{N}) \subset E(\mathbb{Z}) \subset E(\mathbb{Q}) \subset E(\mathbb{R}) \subset E(\mathbb{C}). \tag{9}$$

Puede pues suceder que algunos de ellos sean vacíos, y que al ampliar gradualmente el conjunto de soluciones permitidas, nos encontremos con más y más soluciones. O que al contrario, que todos los conjuntos representados en (9) sean iguales.

Un ejemplo de ecuación diofántica *sin* soluciones reales es la del medio de la tabla (7). En efecto, empleando la notación sugerida en (8), para $E : x^2+y^2+1 = 0$ tenemos

$$E(\mathbb{R}) = \{(x, y) \in \mathbb{R}^2 \text{ tales que } x^2 + y^2 + 1 = 0\} = \emptyset,$$

ya que un número real elevado al cuadrado es positivo, así que el resultado de la suma $x^2 + y^2 + 1$ es siempre al menos 1 y no puede dar 0.

Paradójicamente, aunque los números reales y complejos parecen más difíciles de entender –sobre todo los segundos, al menos a juzgar por el nombre que les hemos dado–, veremos que son éstos los que provocan menos quebraderos de cabeza en las cuestiones que estudiaremos.

Más concretamente, por citar como ejemplo la cuestión que nos ocupa en este capítulo y sobre la cual daremos cuenta en la sección 3, la conjetura de Birch y Swinnerton-Dyer versa sobre la cantidad de soluciones *racionales* de una ecuación diofántica con coeficientes en $\mathbb{Z}$, aventurando cuál debería ser la respuesta de una manera sorprendente y sutil, aunque la comunidad matemática aún da palos de ciego para confirmar que tal intuición es efectivamente correcta.

En cambio, la estructura interna del conjunto de soluciones *reales* o *complejas* de esa *misma* ecuación diofántica está perfectamente bien entendida y no depara ningún misterio. Y sí, más grande no significa más difícil.

Antes de explicar cuál es la dichosa ecuación diofántica estudiada por los matemáticos británicos Bryan Birch y Peter Swinnerton-Dyer y el porqué de su interés y dificultad, hay algunas cuestiones clave que debemos plantearnos:

- *¿Por qué hay ecuaciones diofánticas fáciles y otras difíciles de resolver?*
- *¿Podemos establecer una graduación aproximada del orden de dificultad?*
- *En esa escala de dificultad, ¿cuáles son las que ya sabemos resolver, cuáles las que parecen estar cerca de nuestros dedos y cuáles las que se antojan inalcanzables?*

No es difícil adivinar que el *número* de variables x, y, z, ... de la ecuación es el factor principal que determina el orden de dificultad del problema. En general, a más variables, más salvajemente difícil es el cálculo de las soluciones de la ecuación.

Aunque claro está que siempre hay ejemplos más o menos ridículos que son la excepción a la regla: por poner alguno, la ecuación

$$x + y + z + t = 0 \tag{10}$$

tiene cuatro variables pero no por ello es difícil de resolver. En el conjunto $\mathbb{N}$ de los números naturales no tiene *ninguna* solución, puesto que la suma de cuatro números positivos nunca puede dar 0. En cambio, en cualquiera de los conjuntos $C = \mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$, tiene infinitas soluciones y además podemos describir con mucha exactitud cómo encontrarlas todas: hay libertad absoluta para escoger el valor de las tres primeras variables x, y, z; una vez tomada esa elección, entonces el valor de t ya está comprometido: debe ser $t = -x - y - z$. Y aquí la observación clave, que no sirve para $\mathbb{N}$, es que para cualquiera de los conjuntos C se cumple que

$$\text{si } x, y, z \in C \quad \text{entonces } t = -x - y - z \in C. \tag{11}$$

Así que cuando decimos que el cálculo de las soluciones de una ecuación es más y más difícil cuantas más variables diferentes hay, nos referimos a *casi cualquier* ecuación, digamos a una cualquiera *tomada entre el montón* como por ejemplo,

$$11x^3 - 3xy^5 + 24z^7 - 12xyt + t^6 - 37 = 0,$$

y no a una preparada a propósito de antemano para que sea fácil de resolver. Esta frase, escrita así, pide a gritos que se aclare su significado preciso y se determine con exactitud qué significan términos tan vagos como *casi cualquier* o *tomada entre*

el montón. Para no perder el hilo no nos adentraremos aquí en esas disquisiciones, pero si sois amantes del rigor os gustará saber que esos conceptos se pueden tratar meticulosamente.

Volviendo al tema que nos ocupa, ser capaz de efectivamente *calcular* las soluciones de una ecuación ya es a menudo pedir mucho! Frecuentemente es incluso dificilísimo saber si *hay* alguna solución o no.

Y si las hay, saber *cuántas* hay, aunque no se sepa calcularlas explícitamente. Por pedir aún menos, a uno le gustaría poder mirar una ecuación y decir: existen infinitas soluciones, o no, sólo hay un número *finito* de ellas.

Nada de eso se sabe hacer (excepto en ejemplos muy favorables) cuando el número de variables es tres o más! Por todos estos motivos, en estas notas no diremos nada sobre ecuaciones diofánticas de más de dos variables.

1.3. Ecuaciones de una variable. Por el contrario, la mayoría de esos interrogantes se saben responder razonablemente bien cuando la ecuación diofántica tiene una sola variable: en este caso la ecuación es

$$E : f(x) = a_d x^d + a_{d-1} x^{d-1} + ... + a_2 x^2 + a_1 x + a_0 = 0, \tag{12}$$

donde $f(x)$ es un polinomio cuyos coeficientes son los números a_0, ..., a_d. Supongamos que $a_d \neq 0$, de manera que el grado de f es d.

Teorema 1.1 (Teorema Fundamental del Álgebra)**.** *La ecuación $f(x) = 0$ tiene exactamente d soluciones, si contamos cada una de ellas con su multiplicidad. Concretamente, el polinomio $f(x)$ se puede descomponer como*

$$f(x) = (x - \alpha_1) \cdot ... \cdot (x - \alpha_d)$$

de manera única salvo el orden de disposición de los factores, donde α_1, ..., $\alpha_d \in \mathbb{C}$ son números complejos, no necesariamente todos diferentes. Estos d números son las únicas soluciones de la ecuación y también se conocen como las raíces *del polinomio $f(x)$.*

El teorema no nos da ninguna receta para *calcular* explícitamente cuáles son esas d soluciones, pero al menos nos dice cuántas hay exactamente. Cuando $d \leq 4$, se disponen de fórmulas explícitas para su cálculo a partir de los coeficientes a_0, ..., a_d.

Explícitamente, para $d = 1$, la solución de

$$f(x) = a_1 x + a_0 = 0$$

es simplemente $x = -\frac{a_0}{a_1}$.

Para $d = 2$, las dos raíces del polinomio

$$f(x) = a_2 x^2 + a_1 x + a_0 = 0$$

son las que uno puede calcular mediante la fórmula que todos aprendemos de pequeños:

$$x = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0 a_2}}{-2a_2}. \tag{13}$$

Cuando $d = 3$ o 4, se conocen también fórmulas similares, un poco más enrevesadas pero al fin y al cabo fáciles de aplicar y disponibles en cualquier *software* de cálculo simbólico.

Para $d \geq 5$, en cambio, en general *no existen* fórmulas que permitan calcular las soluciones de (12) como una expresión de sumas, restas, multiplicaciones, divisiones y extracción de raíces de los coeficientes a_0, ..., a_d. Y la frase debe interpretarse tal como la he escrito: no es que *no se sepa* cómo hacerlo, sino que se sabe que *no se puede hacer* para un polinomio arbitrario de grado $d \geq 5$. Por poner un ejemplo sencillo, la ecuación

$$E : f(x) = x^5 - x - 1 = 0$$

es uno de esos casos en que tal fórmula no existe. La teoría que profundiza en estas cuestiones recibió su primer impulso por parte del matemático francés Évariste Galois a principios del siglo XIX y hoy en día sigue siendo pieza angular en muchas de las grandes cuestiones que ocupan a la comunidad matemática.

Évariste Galois.

Si los coeficientes a_0, ..., a_d son enteros o racionales y uno sólo está interesado en las soluciones enteras o racionales, hay un método sencillo para calcularlas: primero uno multiplica el polinomio por el mínimo común múltiple de los denominadores de los coeficientes, asegurándose así que el polinomio tiene sus coeficientes en $\mathbb{Z}$, sin cambiar sus raíces. En segundo lugar, se deshace de todas las raíces $x = 0$ que el polinomio pueda tener dividiéndolo por x tantas veces como sea necesario hasta que $a_0 \neq 0$. Finalmente, uno calcula el resto de las soluciones racionales mediante el siguiente sencillo criterio: toda solución racional $x = \frac{m}{n}$ de la ecuación satisface que, cuando la expresamos en su fracción irreducible, m divide a a_0 y n divide a a_d. Podéis practicar este método con ejemplos concretos: ¿cuáles son todas las soluciones *racionales* de la ecuación

$$77 + 81x + 14x^2 + 20x^3 + 9x^4 - 78x^5 - 4x^6 - 10x^7 - 10x^8 + x^9 = 0,$$

si es que hay alguna?

1.4. Ecuaciones de dos variables. Finalmente, llegamos a la cuestión que realmente nos va a tener entretenidos durante el resto del capítulo: ¿cómo está el panorama cuando la ecuación tiene *dos variables*?

Éste es el caso en el que se enmarca la conjetura de Birch y Swinnerton-Dyer, formulada en los años sesenta. Y es un contexto en que los interrogantes básicos planteados más arriba aún no están completamente resueltos y son centro de atención de muchos investigadores que se dedican a la rama de las matemáticas conocida como *teoría de números*.

En efecto, mientras que las ecuaciones diofánticas de una variable se consideran bien entendidas, las de tres o más parecen inexpugnables para la gran mayoría de matemáticos. Así, las ecuaciones diofánticas de dos variables configuran un estadio de dificultad intermedio en el que uno puede atreverse a resolver alguna de las incógnitas.

Dicho esto, quizás os parecerá curioso saber que Peter Swinnerton-Dyer, de 83 años de edad en el momento de escribir estas líneas, dejó de investigar hace años sobre la conjetura que lleva su nombre y sus trabajos están enfocados a las ecuaciones diofánticas de *tres* variables.

Consideremos una ecuación diofántica

$$E : f(x) = 0 \tag{14}$$

de grado d en dos variables. Para el que le guste entender las cosas con ejemplos, ahí van algunos escogidos más o menos entre el montón:

(15)

$E : f = 0$	Grado
$x^2 - 23xy + 7y^2 + 11x - 1 = 0$	2
$-x^3 + y^2 - x - 1 = 0$	3
$3x^3 + 4y^3 + 2x^2 - 3xy + 13y^2 + x - 5 = 0$	3
$-x^4 + 4xy^3 - 5x^2y^2 + 23x^3 - 2xy + y - 10 = 0$	4

Y para el que prefiera ver las cosas escritas de manera completa y abstracta, en este texto hemos acordado que por ecuación diofántica de grado d en dos variables entendemos una ecuación de la forma

$$\begin{array}{rc} E : f(x) = & a_{d,0}x^d + a_{d-1,1}x^{d-1}y + ... + a_{0,d}y^d + \\ & ... + a_{d-1,0}x^{d-1} + a_{d-2,1}x^{d-2}y + \\ & ... + a_{1,0}x + a_{0,1}y + a_{0,0} \ = \ 0 \end{array} \tag{16}$$

donde a_{ij} son números, los coeficientes de la ecuación.

Con el fin de no complicar innecesariamente la discusión, vamos a suponer además que la ecuación escogida no tiene ninguna *singularidad*. Geométricamente, eso significa que en la representación gráfica de la ecuación no hay ningún punto en que la curva se corta a sí misma, o donde uno se podría pinchar. Algebraica y más rigurosamente, una solución (x, y) de la ecuación (16) se llama *singular* si

además es solución de las ecuaciones

$$\begin{cases} \frac{\partial}{\partial x} f(x,y) = 0 \\ \frac{\partial}{\partial y} f(x,y) = 0, \end{cases}$$

donde $\frac{\partial}{\partial x} f(x,y)$ es la derivada del polinomio $f(x,y)$ respecto la variable x (tomando la variable y como constante) y de forma similar $\frac{\partial}{\partial y} f(x,y)$ es la derivada del polinomio $f(x,y)$ respecto la variable y, tomando la variable x como constante.

Cuando decimos que E no tiene ninguna *singularidad*, queremos decir que ninguna de las soluciones (x,y) de E es también simultáneamente solución de $\frac{\partial}{\partial x} f(x,y) = 0$ y $\frac{\partial}{\partial y} f(x,y) = 0$.

Es bastante habitual que una ecuación no tenga ningún punto singular, y fácil de comprobar: veamos por ejemplo que la ecuación

$$f(x,y) = x^3 - y^2 - x = 0 \tag{17}$$

no tiene ninguna singularidad. Calculamos sus derivadas parciales y planteamos el sistema

$$\begin{cases} f(x,y) = x^3 - y^2 - x = 0 \\ \frac{\partial}{\partial x} f(x,y) = 3x^2 - 1 = 0 \\ \frac{\partial}{\partial y} f(x,y) = -2y = 0. \end{cases}$$

Para que (x,y) sea solución de las tres ecuaciones, la tercera nos indica que necesariamente debemos tener $y = 0$, mientras que la segunda nos dice que $x = \pm\sqrt{\frac{1}{3}}$. Pero ni $(\sqrt{\frac{1}{3}}, 0)$ ni $(-\sqrt{\frac{1}{3}}, 0)$ es solución de la primera ecuación. En la figura 3 hemos representado gráficamente las soluciones reales de (17), donde puede observarse la interpretación geométrica de este hecho: la curva no pasa en ningún momento dos veces por el mismo lugar, ni pincha en ninguna esquina: los geómetras dicen que es una curva *suave* en todos sus puntos.

1.5. Soluciones reales y complejas.

Los conjuntos

$$E(\mathbb{R}) = \{(x,y) \in \mathbb{R}^2 \text{ tales que } f(x,y) = 0\}$$

y

$$E(\mathbb{C}) = \{(x,y) \in \mathbb{C}^2 \text{ tales que } f(x,y) = 0\}$$

de soluciones *reales* o *complejas* son relativamente fáciles de entender (al menos en comparación con el conjunto de soluciones enteras o racionales que trataremos a continuación): la mejor manera de describir estos dos conjuntos es *dibujándolos*.

En el caso de las soluciones reales hay dos posibilidades: $E(\mathbb{R})$ es vacío (en cuyo caso no hay nada que dibujar) o es una *curva* en el plano euclidiano $\mathbb{R}^2$. Es además muy fácil discernir si nos encontramos en una situación u otra (cuestión que dejo para que penséis vosotros mismos). Y cuando $E(\mathbb{R})$ no es vacío, no acostumbra a ser una tarea muy ardua la de representar gráficamente la curva: hoy en día cualquier *software* matemático es capaz de hacerlo. En las figuras 1, 2 y 3 van algunos ejemplos:

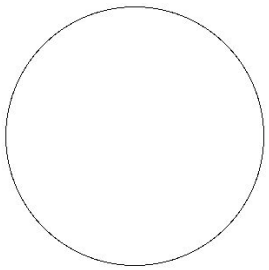


FIGURA 1. Las soluciones reales de $x^2 + y^2 = 1$.

FIGURA 2. Las soluciones reales de $x^2 + y^2 + 1 = 0$. Sí, sí, habéis entendido bien la figura 2: está en blanco porque no hay nada que representar, puesto que la ecuación $x^2 + y^2 + 1 = 0$ no admite soluciones reales.

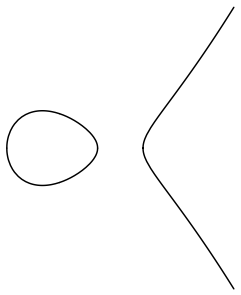


FIGURA 3. Las soluciones reales de $y^2 = x^3 - x$.

En cuanto al conjunto $E(\mathbb{C})$ de soluciones complejas de la ecuación diofántica, el Teorema Fundamental del Álgebra que hemos explicado anteriormente nos muestra que

$$E(\mathbb{C}) \neq \emptyset,$$

es decir, que siempre *existen* soluciones complejas de la ecuación.

En efecto, una manera de entender por qué es la siguiente: escoge un valor $y_1 \in \mathbb{C}$ de la variable y al azar. Cualquiera sirve, mientras el polinomio $f(x, y_1)$ en x que obtenemos al substituir $y = y_1$ no sea constante. Entonces la tarea de encontrar las soluciones (x, y) en $E(\mathbb{C})$ tales que $y = y_1$ no es otra que la de resolver la ecuación

$$f(x, y_1) = 0$$

en la variable x, puesto que ahora y ha dejado de ser una variable para tomar un valor concreto. En estas circunstancias podemos aplicar el Teorema Fundamental del Álgebra: el polinomio $f(x, y_1)$ tiene tantas raíces como su grado (que a lo sumo es d y como mínimo es 1, gracias a la buena elección de y_1 que hemos tomado). Si llamamos x_1 a una de esas raíces, entonces (x_1, y_1) es una solución de E que pertenece al conjunto $E(\mathbb{C})$.

Observad que aunque hayamos dado en escoger y_1 entero, racional o real, no hay ninguna garantía que x_1 sea también entera, racional o real: sólo podemos asegurar que $x_1 \in \mathbb{C}$. Ese hecho, que ya habíamos observado en (11), es crucial para entender por qué a veces es imposible encontrar soluciones enteras, racionales o reales de una ecuación diofántica.

Ejemplifiquemos todo el argumento con una ecuación concreta y sencilla:

$$E : f(x, y) = x^2 + y^2 + 1 = 0.$$

Para mostrar que existen soluciones complejas de E, toma cualquier valor de la variable y: pongamos por caso $y = 5$, aunque todos sirven por igual. Ahora nos preguntamos: ¿para qué valores de x, el par $(x, 5)$ es solución de E? Como 5 es un número entero, algún optimista podría incluso soñar con encontrar algún valor *entero* de x tal que $(x, 5)$ es solución de E, de manera que incluso tendríamos una solución $(x, 5)$ en $E(\mathbb{Z})$. En este ejemplo se da la circunstancia que no se puede encontrar ningún número entero, ni racional, ni real x tal que $(x, 5)$ sea solución de E. La razón es simple: si seguimos el argumento detallado anteriormente, x sería una raíz del polinomio $f(x, 5) = x^2 + 5^2 + 1 = x^2 + 26$. Pero la ecuación de una variable

$$x^2 + 26 = 0 \tag{18}$$

se resuelve sin problemas: las soluciones son los números $x = \pm\sqrt{-26}$, que como todos sabemos *no son reales*, puesto que el cuadrado de todo número real da siempre positivo y no puede dar -26. Así que en este ejemplo

$$E(\mathbb{N}) = E(\mathbb{Z}) = E(\mathbb{Q}) = E(\mathbb{R}) = \emptyset$$

mientras que $E(\mathbb{C})$ es un conjunto infinito que, siguiendo los pasos dados para $y_1 = 5$, sabemos describir con mucha precisión:

$$E(\mathbb{C}) = \{(\pm\sqrt{-1 - y_0^2}, y_1), \quad y_1 \in \mathbb{C}\}.$$

Teniendo en cuenta que $\pm\sqrt{-1 - y_0^2}$ da pie a exactamente dos números complejos diferentes para toda elección de y_1 (excepto para $y_1 = 1$ ó -1), parece natural afirmar que *más o menos hay tantas soluciones complejas de E, como el doble*

de números complejos. ¿No? Y como $2 \cdot \infty = \infty$, pues $E(\mathbb{C})$ es *exactamente tan grande como* el conjunto $\mathbb{C}$. De nuevo esas frases requerirían una justificación más rigurosa, y os animo a buscarla.

Mientras, nosotros nos quedamos satisfechos con este ejemplo de ecuación y la descripción de sus soluciones complejas. Especialmente porque el razonamiento es perfectamente válido para cualquier otra ecuación de dos variables sin singularidades, con lo que tenemos una imagen bastante nítida de la pinta que tiene su conjunto $E(\mathbb{C})$ de soluciones complejas.

Para concluir este apartado, ahí van las representaciones gráficas del conjunto de soluciones complejas de las mismas tres ecuaciones que dibujamos anteriormente:

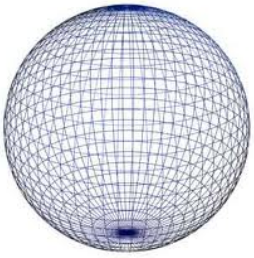


FIGURA 4. Las soluciones complejas de $x^2 + y^2 = 1$.

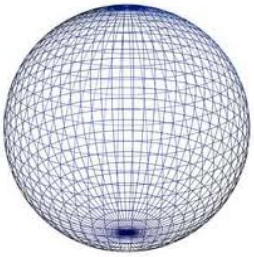


FIGURA 5. Las soluciones complejas de $x^2 + y^2 + 1 = 0$. Sí, sí, también habéis entendido bien esta figura: aunque las ecuaciones de las figuras 4 y 5 no son las mismas, sus conjuntos de soluciones complejas sí son iguales.

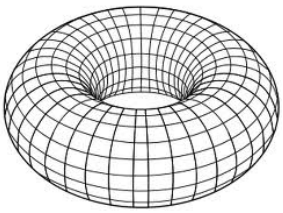


FIGURA 6. Las soluciones complejas de $y^2 = x^3 - x$.

1.6. Soluciones racionales. Como ya había avanzado, aunque los conjuntos $\mathbb{Z}$ de los números enteros y $\mathbb{Q}$ de los números racionales nos son más familiares que los conjuntos $\mathbb{R}$ y $\mathbb{C}$ de los números reales y complejos, calcular o describir los conjuntos

$$E(\mathbb{Z}) \subset E(\mathbb{Q})$$

de soluciones enteras o racionales de una ecuación diofántica E es una tarea mucho más complicada, aunque no lo parezca a simple vista, que la de describir $E(\mathbb{R})$ o $E(\mathbb{C})$ como hemos hecho en el apartado anterior.

Los siguientes enunciados recogen el conocimiento (o mejor dicho, ignorancia) que tenemos actualmente del tema.

(I) Sobre la existencia de soluciones:

Sea $E : f(x, y) = 0$ una ecuación diofántica de dos variables, no-singular, de grado $d \geq 1$ y con coeficientes en el conjunto $\mathbb{Q}$.

(a) Si $d = 2$, conocemos un algoritmo que nos permite decidir si existe alguna solución *racional* de la ecuación E o no, en un número finito de pasos.

(b) Si $d \geq 3$, conocemos un algoritmo que nos permite decidir si existe alguna solución *racional* de la ecuación E o no, y creemos que siempre se puede llevar a cabo en un número finito de pasos, aunque nunca nadie hasta hoy ha podido confirmarlo.

(II) Sobre la cantidad de soluciones:

Supongamos que E admite al menos una solución racional.

(a) Si $d \leq 2$, entonces E tiene infinitas soluciones racionales.

(b) Si $d = 3$, E puede tener un número finito o infinito de soluciones racionales.

(c) Si $d \geq 4$, E tiene sólo un número finito de soluciones racionales.

Vistas las cosas, hay al menos dos cuestiones básicas de las cuales aún no conocemos respuesta.

La primera es la planteada en (Ib). Aunque no lo trataremos aquí, los matemáticos Bjorn Poonen, Alexei Skorobogatov y otros proponen un algoritmo, basado en la llamada obstrucción de Brauer-Manin, para decidir si una ecuación de grado $d \geq 3$ admite alguna solución *racional* o no. Funciona en la práctica con cualquier ejemplo que se ha intentado, pero se ignora aún si funciona realmente para *todas* las ecuaciones. Dicho de otra manera: aunque se intuye que dicho método debería funcionar siempre, no se ha podido excluir nunca la posibilidad de que exista alguna ecuación maldita de grado $d \geq 3$ para la que el algoritmo propuesto no termine nunca.

La segunda cuestión sobre la cual todavía no conocemos respuesta es la planteada en (IIb): si nos dan una ecuación de grado 3, ¿cómo podemos saber si tiene un número finito o infinito de soluciones racionales? Como veremos en las siguientes secciones, la conjetura de Birch y Swinnerton-Dyer proporcionaría, en caso que se pudiera demostrar que es cierta, un criterio explícito para determinar la estructura del conjunto de soluciones racionales de las ecuaciónes diofánticas de dos variables no-singulares de grado 3 con coeficientes racionales.

Y como eso, las *ecuaciónes diofánticas de dos variables no-singulares de grado 3 con coeficientes racionales y con al menos una solución racional* van a pasar a ser nuestro centro de atención en la siguiente sección, pues mejor será que lo abreviemos de alguna forma para no ir repitiendo esa cadena larga de palabras todo el rato. A *esas ecuaciones*, los matemáticos han dado en llamarlas *curvas elípticas*.

Pero que no nos confunda el nombre: su representación gráfica no guarda relación alguna con las elipses. De hecho ya nos encontramos con una curva elíptica en el camino: la ecuación

$$E : y^2 = x^3 - x. \tag{19}$$

En la figura 3 dibujamos el conjunto de sus soluciones reales, donde podemos comprobar que no se parece en nada a una elipse.

En la figura 6 dibujamos el conjunto de sus soluciones complejas; por si no quedo claro entonces, aclarémoslo aquí: el dibujo es un donut, sí, aunque *vacío* por dentro. Es como si de un donut nos hubiéramos comido todo el interior y hubiéramos dejado sólo la *piel*, aunque todos sepamos que es una tontería hablar de la piel de un donut porque nos lo comemos entero.

Qué podemos decir a bote pronto del conjunto de soluciones enteras o racionales de la ecuación (19)? Bueno, lo primero que tenemos que hacer es mirar la ecuación con atención y pensar si vemos a simple vista algún par de valores (x, y) *enteros o racionales* que sean solución. Y de hecho encontrar algunos es muy fácil, no? Por ejemplo, claramente $x = 0$, $y = 0$ satisfacen la ecuación. Y como el número 0 es un número entero (y por tanto también racional), simbólicamente lo ponemos así:

$$(0, 0) \in E(\mathbb{Z}) \subset E(\mathbb{Q}),$$

puesto que, como en otras ocasiones, $E(\mathbb{Z})$ y $E(\mathbb{Q})$ denotan el conjunto de soluciones enteras y racionales de la ecuación.

¿Qué otras soluciones enteras o racionales podéis encontrar a ojo? Yo ahora veo las soluciones $(1,0)$ y $(-1,0)$, que también cumplen la ecuación (19). ¿Veis vosotros otras? Fijaos en que cuando $x < -1$, el valor de $x^3 - x$ es negativo y no puede ser igual al cuadrado y^2 de un número racional, que siempre es positivo.

Así que debemos buscar las soluciones entre los números racionales $x \geq -1$. Si para simplificar el problema nos centramos sólo en encontrar las soluciones *enteras*, ya hemos visto que $x = -1$, $x = 0$ y $x = 1$ dan pie a sendas soluciones enteras $(-1,0)$, $(0,0)$ y $(1,0)$. ¿Podemos hallar otras con $x \geq 2$?

En palabras, ¿estáis de acuerdo que lo que se nos pide es encontrar un número x tal que la diferencia entre su cubo x^3 y él es un número *cuadrado* y^2?

Si hacemos una tabla con los primeros enteros positivos de x y el valor de $x^3 - x$:

x	1	2	3	4	5	6	7	8
$x^3 - x$	0	6	24	60	120	210	336	504

Y paralelamente hacemos una tabla con los primeros enteros $y \geq 0$ y el valor de y^2:

y	0	1	2	3	4	5	6	7	8	9	10	11
y^2	0	1	4	9	16	25	36	49	64	81	100	121
y	12	13	14	15	16	17	18	19	20	21	22	23
y^2	144	169	196	225	256	289	324	361	400	441	484	529

Notad que, hasta donde hemos podido observar, nunca sucede que $y^2 = x^3 - x$ para ninguna combinación de enteros x, y con $x \geq 2$. En la tabla esta afirmación corresponde al hecho que en la fila de valores de $x^3 - x$ no hay ninguno que coincida con ninguno de los valores de y^2, excepto para $x = 1$ e $y = 0$. Parece pues razonable *sospechar* (que es lo que los matemáticos llaman *conjeturar*) que las tres soluciones enteras encontradas son las únicas. Es decir:

Conjetura 1.2. $E(\mathbb{Z}) = \{(-1,0), (0,0), (1,0)\}$.

Quien crea que la sospecha o conjetura es cierta, que se ponga manos a la obra para encontrar un argumento o demostración que muestre que efectivamente lleva razón. Quien crea que la conjetura es falsa, que se ponga a buscar dos números enteros grandes x e y que satisfagan la igualdad $y^2 = x^3 - x$.

¿Se atreve alguien a conjeturar que en realidad

Conjetura 1.3. $E(\mathbb{Q}) = \{(-1,0), (0,0), (1,0)\}$*?*

Quien conteste sí debería reconocer que su afirmación es bastante más atrevida que la anterior, puesto que ello significa que es *imposible* encontrar números racionales $x = \frac{m}{n}$ e $y = \frac{r}{s}$ tales que $\frac{r^2}{s^2} = \frac{m^3}{n^3} - \frac{m}{n}$, y eso ya parece bastante más arduo de comprobar.

Uno no debería confiarse demasiado, pensando que los juicios a primera vista son fiables. Por poner otro ejemplo aparente inocuo, tomad la ecuación

$$E : y^2 = x^3 + 17. \tag{20}$$

¿Podéis encontrar alguna solución entera o racional a simple vista? Resulta que hay exactamente 16 soluciones enteras diferentes, alguna de ellas con valores de x e y bastante grandes, que difícilmente puede ver uno a ojo. Son las siguientes:

$$E(\mathbb{Z}) = \begin{array}{c}\big\{(-1,\pm 4),\ (-2,\pm 3),\ (2,\pm 5),\ (4,\pm 9),\ (8,\pm 23),\\ (43,\pm 282),\ (52,\pm 375),\ (5234,\pm 378661)\big\}.\end{array}$$

Estas son todas las soluciones enteras de (20), pero se puede demostrar que tiene muchas otras soluciones racionales: por ejemplo, podéis comprobar vosotros mismos que

$$P_2 = (137/64, 2651/512),$$
$$P_3 = (298927/40401, 166830380/8120601) \quad \text{y}$$
$$P_4 = (-4531991647/1799117056, -76914444719857/76311349047296)$$

son también soluciones racionales de (20), aunque estaréis de acuerdo que no es fácil encontrarlas a ojo. En realidad, se puede demostrar que en este caso $E(\mathbb{Q})$ es un conjunto *infinito*.

Para concluir, aunque sólo sea para ir familiarizándonos con el término, repitamos la frase del párrafo anterior con esta palabra nueva, *curvas elípticas*, que hemos introducido para abreviar:

La conjetura de Birch y Swinnerton-Dyer proporcionaría, en caso que se pudiera demostrar que es cierta, un criterio explícito para determinar la estructura del conjunto de soluciones racionales de las curvas elípticas.

2. Salir por la tangente

En la sección anterior hemos avanzado ya qué son las curvas elípticas y hemos señalado un motivo (de los muchos que hay: podéis consultar [13], [21], [23], [17] para otros) por el que nos llaman la atención: entre todas las ecuaciones diofánticas no-singulares de dos variables con coeficientes racionales, sólo las de grado $d = 3$ pueden tener un número finito o infinito de soluciones racionales. Y no sabemos cuándo sucede una cosa u otra.

Es más: desearíamos poder describir lo mejor posible el conjunto $E(\mathbb{Q})$ de soluciones racionales de una curva elíptica. En el enunciado (IIa) de la sección anterior, afirmé que las ecuaciones de grado $d = 1$ ó 2 están bajo control: si existe alguna solución racional, sabemos que no hay sólo una sino una infinitud de ellas.

Pero en ese momento no expliqué por qué: hagámoslo ahora, con un ojo puesto en las curvas elípticas y en el motivo por el que el razonamiento que daremos a continuación no es válido cuando el grado de la ecuación diofántica es $d \geq 3$.

2.1. ¿Por qué las ecuaciones de grado $d \leq 2$ tienen infinitas soluciones? Empezemos con las ecuaciones de dos variables de grado 1 con coeficientes racionales, que son las más sencillas. El polinomio que las define es

$$f(x, y) = ax + by + c = 0 \tag{21}$$

donde $a, b, c \in \mathbb{Q}$. Al menos a ó b son diferentes de 0, porque si estos dos coeficientes fueran nulos entonces el polinomio en realidad no tendría grado 1, sino 0.

Estas ecuaciones siempre son no-singulares: $\frac{\partial f}{\partial x} = a$ y $\frac{\partial f}{\partial y} = b$, así que el sistema de ecuaciones

$$\begin{cases} a = 0 \\ b = 0 \end{cases}$$

no tiene solución, por que $a \neq 0$ ó $b \neq 0$.

Y la ecuación (21) siempre tiene soluciones racionales, infinitas de hecho. Para ver por qué, supongamos por ejemplo que $b \neq 0$ (si $b = 0$, entonces $a \neq 0$ y repetimos el mismo razonamiento intercambiando los roles de las variables x e y).

Para cualquier valor $x = x_1$ en el conjunto $\mathbb{Q}$ de los números racionales, obtenemos que

$$y_1 = \frac{-ax_1 - c}{b}$$

tiene sentido precisamente porque $b \neq 0$ y da pie a la solución (x_1, y_1) de la ecuación (21). Además, como $a, b, c, x_1 \in \mathbb{Q}$, concluimos que y_1 también está en el conjunto $\mathbb{Q}$, porque la suma, resta, producto y división de números racionales da un número racional. Y es que $\mathbb{Q}$ es uno de los ejemplos más básicos de lo que la gente llama un *cuerpo*.

La conclusión es que (21) tiene siempre una infinitud de soluciones racionales, *tantas* como números racionales, puesto que hemos sabido asociar una solución racional (x_1, y_1) de (21) a cada $x_1 \in \mathbb{Q}$, y evidentemente la solución (x_1, y_1) determina de manera única el valor x_1.

Sigamos ahora con las ecuaciones de dos variables de grado 2 con coeficientes racionales, que ya no son tan sencillas. La ecuación que las define es

$$E : f(x, y) = ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{22}$$

donde a, b, $c, d, e, f \in \mathbb{Q}$. Una vez más, que la ecuación sea de grado 2 significa que al menos uno de los coeficientes a, b, c es no nulo.

Ahora ya no es verdad que la ecuación (22) sea siempre no-singular. Por ejemplo, cuando $a = c = d = e = f = 0$ y $b = 1$, nos encontramos con la ecuación

$$f(x, y) = x \cdot y = 0,$$

que uno comprueba tiene una singularidad en el punto $(0, 0)$, ya que este punto es efectivamente solución simultánea de las tres ecuaciones $f(x, y) = xy$, $\frac{\partial f}{\partial x} = y$

y $\frac{\partial f}{\partial y} = x$. Ésta es de hecho la única singularidad de la ecuación, que se puede apreciar con claridad en la representación gráfica del conjunto de sus soluciones reales:

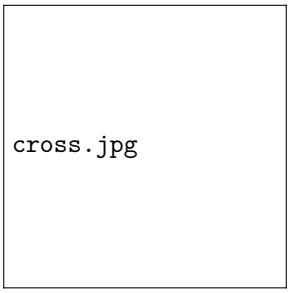


FIGURA 7. Las soluciones reales de $xy = 0$. La singularidad está en $(0, 0)$, el único punto por donde la gráfica *pasa dos veces*.

Mas allá de este ejemplo, en general hay un criterio explícito que nos permite detectar si la ecuación (22) tiene singularidades: eso sucede precisamente cuando la matriz

$$\begin{pmatrix} 2a & b & d \\ b & 2c & e \\ d & e & 2f \end{pmatrix}$$

es *singular*, o, en otras palabras, que su determinante es 0. En el caso de de la ecuación $xy = 0$, por ejemplo, la matriz es

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

que evidentemente tiene determinante 0 y por tanto re-confirma lo que ya sabíamos, que esta ecuación sí tiene singularidades.

Supongamos a continuación que la ecuación (22) es no-singular. A diferencia de las ecuaciones de grado 1, ahora tampoco es verdad que la ecuación (22) siempre tenga alguna solución racional. De hecho ya hemos visto algún ejemplo de eso: en la sección anterior nos encontramos con la ecuación $x^2 + y^2 + 1 = 0$ –que por cierto es no-singular– no tiene ninguna solución racional.

En el enunciado (Ia) comenté que se dispone de un algoritmo que permite decidir si la ecuación (22) tiene o no alguna solución racional en un número finito de pasos. El algoritmo es sencillo y muy fácil de aplicar, pero nos llevaría rato explicarlo. En cambio, estamos especialmente interesados en entender por qué razón si existe una solución, existen infinitas y además podemos describir su conjunto de forma explícita. El motivo es el siguiente:

Tomemos una solución (x_1, y_1) racional cualquiera de la ecuación. Es decir, x_1 e y_1 son números racionales que satisfacen $f(x_1, y_1) = 0$. Consideremos ahora una *recta* que pase por el punto (x_1, y_1), es decir, una ecuación de la forma

$$R_{(r,s)}: \quad r \cdot (x - x_1) + s \cdot (y - y_1) = 0, \tag{23}$$

donde al menos uno de los dos coeficientes, r ó s, es distinto de 0.

Si $s \neq 0$ y poniendo $\lambda = \frac{-r}{s}$, podemos despejar la variable y y obtener la ecuación equivalente

$$R_\lambda: \quad y = \lambda \cdot x + (\, -\lambda x_1 + y_1\,), \tag{24}$$

a la que a lo mejor estamos más acostumbrados. Se trata pues de una recta de *pendiente* λ. Si en cambio $s = 0$, nos queda una ecuación aún más sencilla: $r(x - x_1) = 0$, que, como $r \neq 0$, podemos simplificar y sencillamente quedarnos con

$$x = x_1.$$

En otras palabras, obtenemos la recta *vertical* en el plano con abscisa constante x_1.

¿En cuántos puntos corta la recta $R_{(r,s)}$ el conjunto de soluciones de la ecuación (22)?

En el caso más sencillo en que $s = 0$, la recta tiene ecuación $x = x_1$ y en este caso no estamos preguntando otra cosa que: ¿Cuáles de las soluciones (x, y) de (22) tienen $x = x_1$? Para averiguarlo sólo hace falta hacer la substitución $x = x_1$ en (22). Como la ecuación (22) tiene grado 2 en x e y, el resultado es la ecuación $f(x_1, y)$ de una sola variable –la variable y–, de grado *a lo sumo* 2.

De manera similar, en el caso genérico en que $s \neq 0$, podemos calcular los puntos de intersección de las ecuaciones (23) y (22) haciendo la substitución (24) en (22), obteniendo esta vez un polinomio de grado a lo sumo 2 en la variable x.

En ambos casos insistimos en decir que el polinomio resultante tiene grado a lo sumo 2, porque para algunos valores molestos de r y s puede suceder que dicho polinomio tenga grado 1.

Veámoslo en un ejemplo concreto: la ecuación

$$f(x, y) = xy - 3x - 2y + 5 = 0 \tag{25}$$

es no-singular y pasa por el punto $(x_1, y_1) = (3, 4)$, como se puede comprobar. La intersección con una recta R_λ viene dada por el polinomio

$$\lambda x^2 + (-5\lambda + 1)x + 6\lambda - 3, \tag{26}$$

que siempre tiene grado 2 en x, excepto cuando $\lambda = 0$ que tiene grado 1. Por otro lado, la intersección con la recta vertical $x = 3$ viene dada por el polinomio $y - 4$, que tiene grado 1 en y.

En los casos excepcionales en que el polinomio de la intersección tiene sólo grado 1, la recta $R_{(r,s)}$ corta la curva E descrita por la ecuación (22) en el único punto en que *ya sabíamos* que cortaba: el punto (x_1, y_1) por donde ambas curvas pasan. Cuando esto sucede, diremos que la recta $R_{(r,s)}$ es *excepcional*.

En el caso genérico en que el polinomio de la intersección tiene grado 2, el Teorema Fundamental del Álgebra nos asegura que tiene exactamente *dos* raíces, aunque puede darse el caso que esas dos raíces sean la misma –y decimos que es una raíz de multiplicidad dos–. La traducción geométrica de este hecho es clara: la recta $R_{(r,s)}$ corta la curva E de grado dos en exactamente dos puntos, aunque puede darse el caso que esos dos puntos sean el mismo: eso sucede precisamente cuando la recta $R_{(r,s)}$ es *tangente* a la curva descrita por la ecuación (22).

Veámoslo de nuevo en el ejemplo (25) que tenemos a mano: para calcular explícitamente los dos puntos en que la recta R_λ corta la curva cuando $\lambda \neq 0$, debemos encontrar las dos raíces del polinomio (26), cosa que sabemos hacer sin problemas mediante la fórmula (13):

$$x = \frac{-(-5\lambda+1) \pm \sqrt{(-5\lambda+1)^2 + 4\lambda(-6\lambda+3)}}{2\lambda}$$

$$= \frac{-(-5\lambda+1) \pm \sqrt{\lambda^2+2\lambda+1}}{2\lambda} = \begin{cases} 3 \\ \frac{2\lambda-1}{\lambda}, \end{cases}$$

puesto que $\lambda^2 + 2\lambda + 1 = (\lambda+1)^2$.

Estas dos raíces del polinomio en la variable x proporcionan la primera coordenada de los dos puntos (x, y) de intersección. La coordenada y viene únicamente determinada por x y la ecuación (24).

La primera raíz, $x = 3$, da lugar al punto $(x_1, y_1) = (3, 4)$, que ¡nuevamente sabíamos *de antemano* que debíamos encontrar! La segunda raíz nos da el esperado segundo punto de intersección, que resulta ser –al substituir la raíz en la variable x de (24)–

$$(x_2, y_2) = (\, \frac{2\lambda-1}{\lambda}, -\lambda+3\,)$$

¿Cuándo sucede que $(x_1, y_1) = (x_2, y_2)$? Es decir, ¿Cuándo la recta R_λ es tangente a la curva descrita por la ecuación (25) en el punto $(x_1, y_1) = (3, 4)$?

Bueno, eso es pan comido: $(3, 4) = (\, \frac{2\lambda-1}{\lambda}, -\lambda+3\,)$ precisamente cuando $\lambda = -1$, es decir, cuando tomamos la recta que sale del punto $(3, 4)$ con pendiente -1.

Para el resto de posibles valores del pendiente λ (es decir, cualquier valor excepto $\lambda = 0$, que da lugar a una recta excepcional, y -1, que da pie a la recta tangente), la recta R_λ corta en dos puntos diferentes, (x_1, y_1) y (x_2, y_2), la curva (25).

Si el pendiente λ de la recta es racional, entonces (x_2, y_2) es también una solución racional. Esta afirmación, que no requiere demasiada justificación por obvia (está claro que si $\lambda \in \mathbb{Q}$, entonces $x_2 =, \frac{2\lambda-1}{\lambda} \in \mathbb{Q}$ y $y_2 = -\lambda + 3 \in \mathbb{Q}$ ya que simplemente estamos sumando, restando, multiplicando y dividiendo números racionales), debe ser recordada, porque dejará de ser evidente en contextos futuros.

Este procedimiento, que hemos ilustrado en detalle para el ejemplo (25), es válido para cualquiera otra ecuación (22) que admita como mínimo una solución racional (x_1, y_1), que tomamos como *punto base* de nuestra construcción:

A cada recta $R_{(r,s)}$ no-excepcional con valores (r,s) *racionales*, le sabemos asociar una solución *racional* (x_2, y_2) de (22). Dicha asignación, aunque en principio incurre en dificultades cuando la recta es excepcional, no conlleva ningún problema cuando la recta es tangente a la curva en (x_1, y_1): puesto que en ese caso (x_1, y_1) es una solución doble de la intersección de la recta con la curva, es natural asignar a esta curva el punto (x_1, y_1) mismo. ¿Estáis de acuerdo?

Resumiendo, si denotamos por $\mathbb{Q}^2 \setminus \{0,0\}$ al conjunto de pares (r,s) de números racionales distintos del $(0,0)$, hemos construido una aplicación

$$\left\{ (r,s) \in \mathbb{Q}^2 \setminus \{0,0\},\ \ R_{(r,s)} \text{ no excepcional} \right\} \longrightarrow E(\mathbb{Q}) \tag{27}$$

entre el conjunto de rectas $R_{r,s}$ no-excepcionales de pendiente racional que pasan por el punto base (x_1, y_1) y el conjunto de soluciones racionales de la ecuación E no-singular de grado 2. En el ejemplo (25), los pares (r,s) excepcionales ocurren cuando

$$\begin{cases} r = 0 \ \text{(y } \ s \ \text{ es cualquier número racional no nulo), en cuyo caso } \ \lambda = \frac{0}{s} = 0, \ \text{ ó} \\ s = 0 \ \text{(y } \ r \neq 0 \ \text{ es cualquiera), en cuyo caso } \lambda = \frac{-r}{0} = \infty. \end{cases}$$

En general, se puede comprobar que el número de rectas excepcionales es como mucho 2, como en el ejemplo. Y ese número tiene una bonita interpretación geométrica: el conjunto de las soluciones reales es

- Una elipse, si hay 0 rectas excepcionales de pendiente real;
- Una parábola, si hay 1 recta excepcional de pendiente real;
- Una hipérbole, si hay 2 rectas excepcionales de pendiente real.

¡Comprobadlo vosotros mismos, aunque sólo sea con varios ejemplos tomados al azar!

La función descrita en (27) da cuenta de *todos* los puntos de $E(\mathbb{Q})$, es decir, toda solución racional (x_2, y_2) de (22) se puede obtener mediante esta construcción. En efecto, si $(x_2, y_2) = (x_1, y_1)$, podemos tomar en el conjunto de la izquierda el valor (r_0, s_0) que da lugar a la recta tangente a E en el punto (x_1, y_1). Si $(x_2, y_2$ es distinta de (x_1, y_1), entonces trazamos la recta que une los dos puntos, que de nuevo tendrá pendiente racional y por tanto corresponderá a algún valor $(r,s) \in \mathbb{Q}^2 \setminus \{0,0\}$. En otras palabras, la función es lo que se acostumbra a llamar *exhaustiva*.

Por otro lado, como a lo mejor ya habréis observado, no es cierto que cada valor (r,s) del conjunto de la izquierda proporcione una solución (x_2, y_2) distinta en $E(\mathbb{Q})$. Concretamente, dos pares (r,s) y $(r',s') \in \mathbb{Q}^2 \setminus \{0,0\}$ generan la misma recta $R_{(r,s)} = R_{(r',s')}$ (y por tanto dan lugar a la misma solución (x_2, y_2)) si, y sólo si, $(r',s') = (\lambda r, \lambda s)$ para algún número $\lambda \neq 0$, que es lo mismo que decir que

$$\frac{r}{s} = \frac{r'}{s'},$$

entendiendo que, cuando el denominador es 0, entonces simplemente ponemos $\frac{r}{0} = \infty$.

Así las cosas, vemos que hay *tantas* rectas por (x_1, y_1) como posibles valores diferentes puede tomar el pendiente λ. Y como no tenemos ningún motivo para excluir la recta vertical $x = x_1$ de pendiente infinito, el conjunto de rectas por (x_1, y_1) está en correspondencia natural con el conjunto

$$\mathbb{P}^1(\mathbb{Q}) := \mathbb{Q} \cup \{\infty\}.$$

Este símbolo, $\mathbb{P}^1(\mathbb{Q})$, no es necesario para entender lo que sigue, pero lo hemos introducido porque es una notación habitual en geometría: es lo que se llama la *recta proyectiva* sobre $\mathbb{Q}$. Pero lo único que nos importa aquí es que $\mathbb{P}^1(\mathbb{Q})$ es el conjunto de rectas que pasan por (x_1, y_1) con pendiente λ racional. Y eso nos importa porque una manera un poco más delicada de describir la aplicación (27), si estáis de acuerdo con la discusión anterior, es

$$\mathbb{P}^1(\mathbb{Q}) \setminus \{ \text{ rectas excepcionales } \} \longrightarrow E(\mathbb{Q}), \tag{28}$$

proporcionando una *biyección* o correspondencia uno-a-uno entre

$$\mathbb{P}^1(\mathbb{Q}) \setminus \{ \text{ rectas excepcionales } \}$$

y el conjunto $E(\mathbb{Q})$ de soluciones racionales de la ecuación (22).

A más de uno es posible que le moleste la presencia de esas *rectas excepcionales*, correspondientes a como mucho un par de valores $\lambda_1, \lambda_2 \in \mathbb{P}^1(\mathbb{Q})$, y preferiría una aplicación más limpia y elegante que tratara por igual a todos las rectas que pasan por (x_1, y_1) sin prestar atención a las excepcionales, obteniendo así una aplicación $\mathbb{P}^1(\mathbb{Q}) \stackrel{?}{\longrightarrow} E(\mathbb{Q})$.

Ese es el caso cuando E es una elipse, porque no hay rectas excepcionales. Pero en general hay en efecto una manera uniforme de tratar a todas las curvas no-singulares de grado 2 descritas por (22), independientemente de si son elipses, hipérbolas o parábolas.

Aunque no será imprescindible para entender las siguientes secciones, uno puede cambiar de punto de vista y decir –definir– que $\mathbb{P}^1(\mathbb{Q})$ *es* el conjunto $\bar{E}(\mathbb{Q})$ de los *puntos racionales proyectivos* de la curva E.

Esta definición es bastante natural, en el fondo, como mínimo para los elementos $\lambda \in \mathbb{P}^1(\mathbb{Q})$ que no son excepcionales, puesto que para éstos efectivamente tenemos que λ corresponde, vía (28), a un punto o solución racional $(x, y) \in E(\mathbb{Q})$. Al considerar $\bar{E}(\mathbb{Q})$, aunque parezca artificial la primera vez, tan sólo estamos *ampliando* el conjunto $E(\mathbb{Q})$ con a lo sumo un par de elementos más, los que corresponden a las rectas excepcionales.

Según esta nomenclatura, $E(\mathbb{Q})$ es el conjunto de puntos o soluciones racionales de E, mientras que hemos dado en llamar $\bar{E}(\mathbb{Q})$ el conjunto de puntos racionales proyectivos de E. A los puntos (dos, como mucho) que pertenecen a $\bar{E}(\mathbb{Q})$ pero no a $E(\mathbb{Q})$, es natural llamarlos puntos excepcionales, aunque los geómetras han

acuñado otro término, más sugestivo, y completamente establecido ya en el argot matemático: *puntos del infinito*.

2.2. ¿Qué falla en el razonamiento anterior cuando la ecuación tiene grado $d \geq 3$? Muchas de las ideas que hemos explicado son perfectamente válidas para cualquier ecuación diofántica no-singular de dos variables de grado d cualquiera. Por otro lado, recordad que en el enunciado (IIc) afirmé –sin explicar el porqué– que, cuando $d \geq 4$, existen a lo sumo un número finito de soluciones racionales, lo cual nos indica que parte del argumento anterior *tiene que ser* falso cuando $d \geq 4$.

Emparedadas entre las ecuaciones de grado $d \leq 2$ y las de grado $d \geq 4$ se encuentran las curvas elípticas: las de grado $d = 3$.

Como afirmé en (IIb), existen curvas elípticas que tienen una infinitud de soluciones racionales –éste es el caso de la curva elíptica (20)– y existen otras que tienen tan sólo un número finito de soluciones racionales –éste es el caso de la curva elíptica (19): la respuesta a la pregunta que dejé en el aire en ese momento es que efectivamente las conjeturas A y B son ciertas y tenemos $E(\mathbb{Q}) = \{(-1, 0), (0, 0), (1, 0)\}$; no hay ninguna otra solución racional. ¿Podéis explicar por qué?–.

Será interesante pues averiguar hasta qué punto las ideas expuestas para ecuaciones de grado $d \leq 2$ funcionan sin problemas para ecuaciones de grado $d \geq 3$, y en qué fallan. Como contrapunto, también sería importante entender por qué motivo la demostración de (IIc) sólo sirve para ecuaciones de grado $d \geq 4$ y deja de ser cierta para ecuaciones de grado menor, pero eso nos llevaría demasiado lejos y requeriría maquinaria demasiado pesada para ser explicada de manera clara y breve aquí: el resultado de finitud (IIc) no se descubrió hasta 1983 en [6] y se debe al matemático alemán Gerd Faltings, con el que ganó la prestigiosa *medalla Fields*.

Consideremos pues una ecuación no-singular

$$E: \quad f(x, y) = 0 \tag{29}$$

de grado $d \geq 3$ y con coeficientes racionales. Queremos rehacer los pasos que hicimos con anterioridad cuando el grado de la ecuación era $d = 2$, donde supusimos que existe al menos una solución racional (x_1, y_1) y acabamos demostrando que existe una infinitud de de ellas.

Supongamos pues también que existe como mínimo una solución racional (x_1, y_1) de (29) y consideremos una recta $R_{(r,s)}$ cualquiera de ecuación (23) que pase por el punto (x_1, y_1). ¿Cuántas soluciones en común tienen las ecuaciones (23) y (29)? Equivalentemente, en su formulación geométrica: ¿En cuántos puntos del plano se cortan la recta $R_{(r,s)}$ con la curva E?

El razonamiento es muy similar al que ya hicimos: para calcular esta intersección sencillamente tenemos que resolver el sistema de ecuaciones dado por (23) y (29). En la práctica, consiste en aislar la variable x o y –como hicimos en (24)– y substituirla en la ecuación (29), obteniendo un polinomio de una variable de grado *a lo sumo* d.

Como entonces, salvo unas cuantas rectas excepcionales, el polinomio de la intersección tiene exactamente grado d; descartemos por el momento esas rectas excepcionales y supongamos que éste es el caso.

Por el Teorema Fundamental del Álgebra, el polinomio tiene precisamente d raíces, que dan lugar a d soluciones

$$(x_1, y_1),\ (x_2, y_2), ...,\ (x_d, y_d),$$

de las cuales una de ellas ya sabemos que debe ser la solución base (x_1, y_1).

Como antes, nadie nos asegura que estas d raíces sean diferentes: puede haber algunas con multiplicidad doble, triple, etc, y eso conlleva la posibilidad que algunas de las soluciones (x_i, y_i) estén repetidas.

Todo quedará mucho más claro si lo ilustramos con un ejemplo. Retomemos la curva elíptica (19) que ya apareció anteriormente y que tiene ecuación de grado $d = 3$:

$$E : y^2 = x^3 - x. \tag{30}$$

Podemos tomar como punto base racional cualquiera de los tres puntos que ya encontramos, por ejemplo $(x_1, y_1) = (0, 0)$. El haz de rectas que pasa por este punto está formado por las rectas de ecuación $R_{(r,0)} : x = 0$ ó

$$R_\lambda : \quad y = \lambda \cdot x, \tag{31}$$

que es lo que obtenemos al substituir $x_1 = 0$, $y_1 = 0$ en (24).

El polinomio de la intersección de la recta $x = 0$ con E se obtiene substituyendo $x = 0$ en (30): da $y^2 = 0$, que tiene $y = 0$ como única *raíz doble*. En este caso la recta $x = 0$ es *excepcional*, pues corta la curva elíptica tan sólo en el punto $(x_1, y_1) = (0, 0)$, contado con multiplicidad *dos*.

Calculemos a continuación la intersección de cualquiera de las rectas (31) con (30): el polinomio en la variable x que obtenemos al substituir la primera ecuación en la segunda es

$$x^3 - \lambda^2 x^2 - x = x(x^2 - \lambda^2 x - 1) = 0, \tag{32}$$

que tiene como raíces $x_1 = 0$ (que ya esperábamos) y

$$x_2, x_3 = \frac{\lambda^2 \pm \sqrt{\lambda^4 + 4}}{2},$$

las dos raíces del polinomio de segundo grado $x^2 - \lambda^2 x - 1$, una para cada signo + y −. ¿Se os ocurre alguna manera de simplificar el término $\sqrt{\lambda^4 + 4}$? A mí no, así que lo dejo así, aunque recordad que en el ejemplo anterior que resolvimos también con todo detalle, nos encontramos con $\sqrt{\lambda^2 + 2\lambda + 1}$, que sí supimos simplificar.

Comoquiera que sea, a cada una de las tres raíces x_1, x_2, x_3 le corresponde una solución de (30), concretamente:

$$(x_1, y_1) = (0, 0), \quad (x_2, y_2) = \left(\frac{\lambda^2 + \sqrt{\lambda^4 + 4}}{2}, \frac{\lambda^3 + \lambda\sqrt{\lambda^4 + 4}}{2}\right) \quad \text{y} \tag{33}$$

$$(x_3, y_3) = \left(\frac{\lambda^2 - \sqrt{\lambda^4 + 4}}{2}, \frac{\lambda^3 - \lambda\sqrt{\lambda^4 + 4}}{2}\right). \tag{34}$$

Hasta aquí hemos procedido exactamente igual que antes. Siendo optimistas, podríamos decir que hasta nos ha ido mejor en esta ocasión: partiendo del punto base (x_1, y_1), a cada recta R_λ le podemos asignar no una, sino dos soluciones nuevas de (30): (x_2, y_2) y (x_3, y_3).

El problema está en que nosotros estamos interesados en encontrar soluciones *racionales*. Claro está que $(x_1, y_1) = (0, 0)$ tiene sus coordenadas racionales (¡de hecho enteras!), pero aunque tomemos valores *racionales* del pendiente λ, ¿podemos asegurar que (x_2, y_2) ó (x_3, y_3) son soluciones racionales?

Como ya hemos repetido varias veces, el conjunto $\mathbb{Q}$ es un *cuerpo*: las operaciones de sumar, restar, multiplicar y dividir números racionales tienen como resultado de nuevo un número racional. Pero el cálculo de cualquiera de las coordenadas x_2, y_2, x_3 ó y_3 a partir de λ involucra otra operación más: la extracción de una raíz cuadrada. Y todos sabemos (conozcamos o no el concepto de *cuerpo*, que en el fondo no aporta nada nuevo a la discusión), que la raíz cuadrada de un número racional puede no ser racional. De hecho ¡hasta es lo habitual! Para cualquiera de los valores

$$n = 2, 3, \frac{1}{5}, -7 \in \mathbb{Q}, \text{ por ejemplo, } \sqrt{n} \notin \mathbb{Q}.$$

Éste es precisamente el motivo por el cual el argumento que dimos en el apartado anterior para demostrar que las ecuaciones de grado 2 con al menos una solución racional (x_1, y_1) tienen infinitas soluciones racionales. En ese caso, para *cada* valor de $\lambda \in \mathbb{Q}$ –excepto para los dichosos excepcionales–, obteníamos *una nueva* solución racional en $E(\mathbb{Q})$.

En cambio, cuando el grado de la ecuación es $d \geq 3$, el mismo procedimiento –que hemos explicado en toda generalidad, e ilustrado en detalle para la curva elíptica (30)– nos permite asignar a cada valor $\lambda \in \mathbb{Q}$ –excepto los pocos excepcionales, que como mucho hay d diferentes– *dos nuevas* soluciones, que *raramente* son racionales.

Vayamos con cuidado con las cosas que afirmamos, no vaya a ser que os confunda: esto muestra por qué el argumento dado para $d = 2$ falla para $d \geq 3$. Pero *no* demuestra que una ecuación de grado $d \geq 3$ pueda tener sólo un número finito de soluciones racionales.

De hecho ya conocemos el ejemplo de la ecuación (20) de grado 3 que tiene infinitas soluciones racionales y hemos insistido en que la demostración de Faltings del hecho cierto que las ecuaciones de grado $d \geq 4$ tienen sólo un número finito de soluciones racionales es profunda y difícil.

Ahora que ya entendemos qué falla en el razonamiento cuando la ecuación tiene grado $d \geq 3$, nos preguntamos:

2.3. ¿Hay algo que se pueda aprovechar? Para empezar, las fórmulas explícitas que hemos encontrado en (33) y (34) para las soluciones (x_2, y_2) y (x_3, y_3) de la ecuación (30), ¿no creéis que a lo mejor podrían servirnos para demostrar la Conjetura 1.3? Ya hemos avanzado anteriormente que la conjetura es cierta, pero dejamos como interrogante el explicar por qué.

A continuación va una posible estrategia, que vosotros decidiréis si merece ser proseguida o no. Aunque ya hemos explicado que, dado un valor λ, las coordenadas de (x_2, y_2) y (x_3, y_3) pueden ser no racionales, planteemos ahora la pregunta recíproca: si existe una solución racional $(a, b) \in E(\mathbb{Q})$ diferente de $(x_1, y_1) = (0, 0)$, ¿podemos construirla mediante el procedimiento anterior a partir de una recta R_λ con $\lambda \in \mathbb{Q}$?

Estaréis de acuerdo conmigo en que sí: trazáis la recta

$$R_\lambda : y = \frac{b}{a}x,$$

que tiene pendiente *racional* $\lambda = \frac{b}{a} \in \mathbb{Q}$ porque $a, b \in \mathbb{Q}$ y $a \neq 0$ por hipótesis, y une el punto $(0, 0)$ con el punto (a, b), como se puede comprobar. Así que, por construcción, si rehacemos los cálculos a partir de esta recta, encontraremos (a, b) como una de las dos nuevas soluciones (x_2, y_2) ó (x_3, y_3).

La conclusión es que si queremos calcular el conjunto $E(\mathbb{Q})$ total de soluciones racionales de (30), podemos proceder así: supongamos que $(a, b) \in E(\mathbb{Q})$ es una de esas soluciones. Debe por tanto corresponder a un pendiente racional $\lambda \in \mathbb{Q}$. Las fórmulas (33) y (34) nos permiten recuperar de nuevo el punto (a, b) a partir del valor λ, y deducimos que (a, b) es racional si y sólo si $\sqrt{\lambda^4 + 4} \in \mathbb{Q}$, es decir, si $\lambda^4 + 4$ es un número *cuadrado* en $\mathbb{Q}$.

Esto efectivamente sucede cuando tomamos $\lambda = 0$, porque $4 = 2^2$. En este caso, la recta R_λ es la recta horizontal del eje de las x, que corta la curva elíptica en el punto base $(0, 0)$ y los dos nuevos puntos $(x_2, y_2) = (-1, 0)$ y $(x_3, y_3) = (1, 0)$. ¡Estos son los puntos que ya sabíamos que existían en $E(\mathbb{Q})$!

Resumiendo, la conjetura 1.3 es equivalente a la siguiente:

Conjetura 2.1. *El único número racional $\lambda \in \mathbb{Q}$ tal que $\lambda^4 + 4$ es un número cuadrado en $\mathbb{Q}$ es $\lambda = 0$.*

Dejo para vosotros el problema de decidir si es una buena estrategia demostrar la veracidad de la conjetura 1.3 mostrando que la conjetura 2.1 es cierta, o no. Comoquiera que sea, hemos traducido la cuestión del cálculo de soluciones racionales en una curva elíptica en un problema aritmético de planteamiento muy sencillo.

Pero más allá de esta ecuación concreta, ¿hay algo que se pueda aprovechar del argumento anterior para efectivamente construir nuevas soluciones *racionales* a partir de una solución racional base (x_1, y_1) en cualquiera ecuación no-singular de grado $d \geq 3$?

Una idea interesante es la siguiente: si en lugar de tomar una recta de pendiente racional cualquiera, tomamos la recta *tangente* R_{λ_0} al punto (x_1, y_1), entonces la solución (x_1, y_1) aparece como mínimo dos veces entre el conjunto

$$(x_1, y_1), (x_2, y_2), ..., (x_d, y_d)$$

de puntos de intersección de la recta con la curva E descrita por la ecuación (29).

Cuando uno hace el cálculo con varios ejemplos –y a continuación desarrollaremos uno en detalle, para ilustrar mejor lo que estoy explicando–, rápidamente se da cuenta que habitualmente la recta R_{λ_0} es no-excepcional y (x_1, y_1) aparece exactamente dos veces y no más entre el conjunto de d puntos de intersección de la recta con la curva (aquellos puntos (x_1, y_1) sobre cuya recta tangente pasa con multiplicidad mayor o igual que 3 se llaman *puntos de inflexión*, y son bastante particulares).

Si esto sucede, y además el grado de la ecuación es $d = 3$, entonces esta construcción arroja una única nueva solución (x_3, y_3) más allá de la solución doble $(x_1, y_1) = (x_2, y_2)$. Lo más relevante de esta situación, y el motivo por el que estamos interesados en ella, es que:

(35) ¡(x_3, y_3) es también una solución racional!

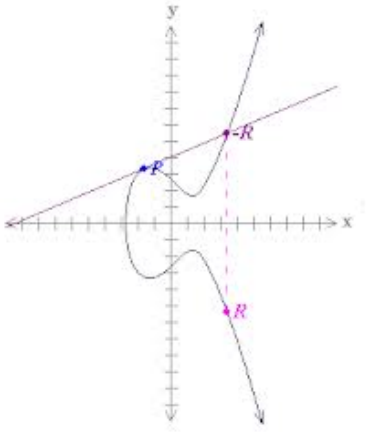


FIGURA 8. La recta tangente al punto P corta la curva elíptica en un tercer punto más.

El porqué no es para nada difícil: la intersección de la recta tangente con la curva E se calcula como lo hemos hecho ya en varias ocasiones: resolviendo el sistema de ecuaciones dado por (24) y (29), que al substituir la primera en la segunda se convierte en un polinomio $p(x)$ de grado 3 en la variable x. De las tres raíces que el Teorema Fundamental del Álgebra nos asegura que tiene, dos de ellas ya sabemos que son la primera coordenada x_1 del punto base, contada dos veces. Así, el polinomio descompone:

$$p(x) = (x - x_1)^2(x - x_3),$$

donde x_3 la primera coordenada de la nueva solución (x_3, y_3).

Para mostrar que $x_3 \in \mathbb{Q}$, podemos observar sencillamente que $(x - x_3) = \frac{p(x)}{x - x_3}$ es el cociente de dos polinomios con coeficientes racionales. Y claro está que ese

cociente tendrá de nuevo coeficientes racionales: en particular, el término independiente x_3 es racional. Finalmente, usando (24), deducimos que $y_3 = \lambda_0 \cdot x + (-\lambda_0 x_1 + y_1)$, que nuevamente es racional porque todos los números involucrados en su cálculo son racionales.

Esta construcción nos proporciona un algoritmo –llamémosle el *algoritmo de la recta tangente*– para encontrar, en condiciones favorables, soluciones racionales a partir de una dada: si empezamos con una solución racional (x_1, y_1), podemos trazar la recta tangente R_{λ_0} al punto y encontrar –si R_{λ_0} no es excepcional y (x_1, y_1) no es un punto de inflexión– una nueva solución racional (x_3, y_3). Podemos ahora tomar (x_3, y_3) y repetir el procedimiento, y así sucesivamente.

Claro está que a veces el algoritmo de la recta tangente se interrumpe, cuando por ejemplo nos encontramos un punto de inflexión. Y otras veces la secuencia de puntos que hayamos es cíclica, produciendo tan sólo una cantidad finita de soluciones racionales diferentes. Esto sucede forzosamente cuando la curva elíptica que hemos tomado tiene un número finito de puntos racionales, lo sepamos o no antes de empezar.

El ejemplo (30) es uno de ellos, como sabemos ya, puesto que hemos afirmado que la conjetura 1.3 es cierta. Si llevamos a cabo el algoritmo tomando como punto base (x_1, y_1) cualquiera de las tres soluciones racionales $(-1, 0)$, $(0, 0)$ ó $(1, 0)$, podemos observar en la figura 3 que la recta tangente a cualquiera de ellos es la recta vertical $x = x_1$ que pasa por la coordenada x_1 del punto. En los tres casos, la recta tangente es excepcional: pasa por (x_1, y_1) con multiplicidad 2 pero no corta a ningún tercer punto (x_3, y_3); usando la terminología descrita al final de la sección anterior, diríamos que $(x_3, y_3) = \infty$ es el punto del infinito. Quien está familiarizado con la geometría proyectiva, quizás pensará: si es verdad que al fin y al cabo podemos tratar el punto del infinito como un punto más, ¿no podemos seguir con el algoritmo tomando este punto como base? En efecto. Pero puede comprobarse que el punto del infinito es un punto de inflexión, así que comoquiera que lo llamemos o lo miremos, el algoritmo de la recta tangente se interrumpe ahí mismo, y no arroja ninguna solución racional más.

En muchas otras ocasiones el algoritmo de la recta tangente se prolonga indefinidamente, sin interrumpirse jamás ni entrar en una sucesión periódica de soluciones racionales. En estos casos, el conjunto $E(\mathbb{Q})$ de puntos o soluciones racionales es *infinito*. Ilustremos el algoritmo de la recta tangente con la curva elíptica (20), para ver en detalle una situación en la que esto sucede.

Recordemos que la ecuación de esa curva elíptica es

$$E: \quad y^2 = x^3 + 17 \tag{36}$$

y que una solución entera es $(x_1, y_1) = (-1, 4)$. Un breve cálculo muestra que la recta tangente a la curva E por este punto base tiene ecuación

$$R: \quad y = \frac{3}{8}x + \frac{35}{8}; \tag{37}$$

es la recta de pendiente racional $\lambda_0 = \frac{3}{8}$ que pasa por el punto.

Para comprobarlo, y de paso calcular el *tercer* punto de intersección de la recta tangente con la curva, substituimos una ecuación en la otra y obtenemos el polinomio de grado 3 en la variable x:

$$p(x) = (\frac{3}{8}x + \frac{35}{8})^2 - x^3 - 17 = \frac{137}{64} + \frac{105}{32}x + \frac{9}{64}x^2 - x^3.$$

Para encontrar las tres raíces del polinomio, es útil aprovechar la información que tenemos a mano: dos de las raíces deberían ser la coordenada $x_1 = -1$ con multiplicidad dos, así que el polinomio $(x+1)^2$ debe dividir exactamente a $p(x)$. Si efectuamos la división, obtenemos

$$\frac{p(x)}{(x+1)^2} = -x + \frac{137}{64}, \quad \text{con lo que} \quad p(x) = (x+1)^2(-x + \frac{137}{64})$$

y por tanto la nueva solución que encontramos es

$$(x_3, y_3) = (x_3, \frac{3}{8}x_3 + \frac{35}{8}) = (\frac{137}{64}, \frac{2651}{512}),$$

que es la que ya llamamos P_2 cuando introdujimos la curva (20) en la sección anterior, y que entonces no había explicado cómo calcularla. Si repetimos el proceso con este punto base, encontramos la solución racional que llamamos antes P_4. En cambio, la solución racional

$$P_3 = (298927/40401, 166830380/8120601)$$

no se encuentra en la órbita de (x_1, y_1). En la siguiente sección explicaré cómo calcularla, porque de hecho daremos con una descripción explícita de *todos* los puntos de $E(\mathbb{Q})$.

Más allá de este ejemplo concreto, es conveniente subrayar que el recíproco de una de las afirmaciones que formulé anteriormente también es cierto:

Teorema 2.2. *Una curva elíptica E tiene infinitos puntos racionales si, y sólo si, existe alguna solución racional (x_1, y_1) a partir de la cual el algoritmo de la recta tangente produce infinitas soluciones racionales diferentes.*

Fijaos en que una de las implicaciones del enunciado es obvia: si el algoritmo de la recta tangente produce una infinitud de soluciones racionales diferentes, está claro que $E(\mathbb{Q})$ es un conjunto infinito. Pero la implicación contraria, que afirmé ser cierta, no es tan evidente: en principio podría pasar que, aunque $E(\mathbb{Q})$ fuera infinito, el algoritmo de la recta tangente se interrumpiera o fuera cíclico sistemáticamente, sin importar qué solución racional base tomáramos.

Este teorema es interesante en cuanto nos indica que el algoritmo de la recta tangente es realmente fructífero: sirve para demostrar que $E(\mathbb{Q})$ es infinito en algunas ocasiones y permite calcular una cantidad infinita de puntos diferentes de este conjunto.

Pero no es verdad que permita calcularlos *todos*: si denotamos $O(x_1, y_1)$ a la órbita del punto base (x_1, y_1), es decir, a la sucesión de puntos racionales en $E(\mathbb{Q})$ que obtenemos a partir de (x_1, y_1) aplicando el algoritmo, se tiene que:

Proposición 2.3. *Sea E una curva elíptica con infinitos puntos racionales. Para cualquier punto racional $(x_1, y_1) \in E(\mathbb{Q})$, hay infinitos puntos racionales en $E(\mathbb{Q})$ que no están en $O(x_1, y_1)$.*

En la siguiente sección dispondremos de las herramientas necesarias para explicar porque el teorema 2.2 y la proposición 2.3 son ciertos: concretamente, véase la discusión que sigue a la proposición 3.3.

Más allá de la limitación del teorema 2.2 puesta de manifiesto en la proposición 2.3, hay un motivo más importante por el que este teorema no es demasiado satisfactorio: el teorema 2.2 simplemente *traduce* la cuestión de si $E(\mathbb{Q})$ es un conjunto finito o infinito, ¡en otra cuestión igualmente difícil de responder!

Me explico: aunque *en teoría* el algoritmo de la recta tangente sirve para detectar si $E(\mathbb{Q})$ es un conjunto finito o infinito, *en la práctica* es de dudosa aplicación.

Si $E(\mathbb{Q})$ es finito pero no lo sabemos y queremos demostrarlo, el teorema 2.2 es en realidad muy poco útil. Aunque comprobemos que la órbita $O(x_1, y_1)$ de cualquier punto racional base (x_1, y_1) que vayamos encontrando es finita, ¿cómo podemos garantizar que no existe algún punto base para el que la órbita es infinita? Imposible con las ideas que hemos desarrollado hasta el momento; la conjetura de Birch y Swinnerton-Dyer, cuya formulación exacta es el objetivo último de este capítulo, nos dará la clave.

Si por el contrario $E(\mathbb{Q})$ es infinito –aunque eso a priori no lo sepamos–, el teorema 2.2 nos da un criterio para comprobar que así es: *existe* un punto base racional (x_1, y_1) tal que $O(x_1, y_1)$ es una órbita infinita; si somos capaces de encontrarlo, esto implicará que en efecto $E(\mathbb{Q})$ es infinito. Pero ¿cómo encontrarlo? ¡No sabemos cómo! Será puramente una cuestión de suerte el dar con tal afortunado punto (x_1, y_1), y no podemos fiarnos de un criterio que dependa de eso.

Vistas las cosas, está claro que debemos profundizar y refinar más nuestros argumentos.

Cuando el grado de la ecuación es mayor, $d \geq 4$, la idea del algoritmo de la recta tangente puede llevarse a término sólo cuando la recta tangente R_{λ_0} pasa por el punto base (x_1, y_1) con multiplicidad $d - 1$. Cuando esto sucede, el *otro* punto de intersección (x_d, y_d) es, por la misma razón que antes, racional. Pero como $d - 1 \geq 3$, (x_1, y_1) es un punto de inflexión y ya hemos comentado que son poco comunes y hay sólo un número finito de ellos. En consecuencia uno no puede esperar construir infinitos puntos racionales diferentes por este procedimiento, lo cual casa perfectamente con el teorema de Faltings.

La aplicabilidad de la idea "partimos de un punto racional base (x_1, y_1) y construimos otros trazando rectas de pendiente racional desde él", que empecé describiendo para las ecuaciones de grado $d = 2$, se puede resumir en el cuadro siguiente:

- Si $d = 2$, cualquier recta de pendiente racional por (x_1, y_1), salvo como mucho dos rectas excepcionales, produce una solución racional nueva. Hay por tanto una infinitud de soluciones racionales diferentes.
- Si $d = 3$, sólo la recta tangente a (x_1, y_1) puede producir una solución racional nueva. Iterando el algoritmo, en ocasiones se encuentra un conjunto infinito de soluciones racionales diferentes, y en otras el algoritmo se interrumpe o es cíclico.
- Si $d \geq 4$, sólo cuando la recta tangente a (x_1, y_1) pasa por el punto con multiplicidad $d - 1$ podemos obtener una nueva solución racional. Y sólo cuando esa nueva solución satisface lo mismo podemos iterar el algoritmo. Pero como mucho podremos llegar a realizar un número finito de iteraciones, porque el conjunto de puntos de inflexión en la curva es finito.

3. Curvas elípticas y la conjetura de Birch y Swinnerton-Dyer

El objetivo de esta última sección es explicar la conjetura de Birch y Swinnerton-Dyer, en la versión propuesta por el Clay Mathematics Institute, cuya resolución permitiría dar una descripción muy satisfactoria de la *estructura algebraica* del conjunto $E(\mathbb{Q})$ de puntos racionales de una curva elíptica, a partir del *orden de anulación* en el punto $s = 1$ de una función analítica

$$L_E(s) : \mathbb{C} \longrightarrow \mathbb{C},$$

la *función zeta* o *función* L de E, como suele llamarse. Existe una versión más refinada de la conjetura, fuera del alcance de estas notas y que por tanto no formularemos, que incluso predice cuál debería ser el valor del primer coeficiente no nulo en el desarrollo de Taylor de la función $L_E(s)$ en $s = 1$ a partir del comportamiento aritmético de E.

En cualquier caso, aunque todavía no haya tenido tiempo de introducir propiamente todos los actores en juego, podemos vislumbrar ya que la conjetura de Birch y Swinnerton-Dyer establece un sorprendente puente entre un objeto puramente algebraico, $E(\mathbb{Q})$, y otro analítico, la función de variable compleja $L_E(s)$ que en breve definiremos.

Antes de hacerlo, recordemos qué entendemos por *curva elíptica*:

Definición 3.1. *Una curva elíptica es una ecuación diofántica*

$$E: \quad f(x, y) = 0$$

de dos variables, no-singular, de grado 3 *y coeficientes racionales, junto con un punto racional base*

$$O \in E(\mathbb{Q}) = \{\, (x, y) \in \mathbb{Q} \times \mathbb{Q},\, f(x, y) = 0 \,\}$$

fijado en el conjunto de soluciones racionales.

Implícita pues en la definición es la hipótesis de que $E(\mathbb{Q})$ no es el conjunto vacío –cosa que, como ya vimos, puede suceder perfectamente. Dada esa hipótesis, es entonces lícito considerar y fijar un punto racional cualquiera en $E(\mathbb{Q})$. Por motivos que quedarán claros en breve, preferimos llamar O al punto base que fijamos, en lugar de (x_1, y_1) como iba haciendo en la sección anterior.

En realidad estamos tratando con una familia muy concreta y explícita de ecuaciones diofánticas: se puede comprobar –mediante el llamado teorema de Riemann-Roch– que toda curva elíptica admite un cambio de variables que la transforma en una ecuación de la forma

$$E: \quad y^2 = x^3 + Ax + B \tag{38}$$

donde $A, B \in \mathbb{Q}$ son números racionales tales que $\Delta(E) := -16(4A^3 + 27B^2) \neq 0$.

Estas ecuaciones simplificadas se llaman *ecuaciones de Weierstrass*, y la última condición es la que asegura que la ecuación es no-singular; el número $\Delta(E)$ se acostumbra a llamar el *discriminante* de la curva elíptica E.

Pero lo más importante de todo es que, al transformar la ecuación de una curva elíptica cualquiera en una ecuación de Weierstrass, el cambio de variables transforma el punto fijado O en el *punto del infinito* de la ecuación (38). Quien conoce los rudimentos de la geometría proyectiva, puede calcular rápidamente a qué punto nos referimos. Pero no hace falta ni mucho menos desarrollar esta teoría aquí, porque nos llevaría demasiado lejos –y que por otro lado recomiendo estudiar de manera entusiasta; un libros donde iniciarse es [18], y otro que cubre la teoría fundamental de curvas algebraicas es [7]–. Basta interpretar el punto O como el punto que *deberíamos encontrar* como tercer punto de intersección de cualquier recta vertical $x = x_0$ con E. En efecto, observad que si hacemos la substitución $x = x_0$ en (38) obtenemos la ecuación en la variable y:

$$y^2 = x_0^3 + Ax_0 + B,$$

que *sólo* tiene grado 2. En la terminología empleada en la sección anterior, las rectas verticales son todas ellas excepcionales, como puede observarse en el ejemplo representado gráficamente en la figura 3: toda recta vertical corta la curva en dos puntos y no en tres.

En completa analogía con el razonamiento que dimos al final del apartado 2.1, *añadimos* un punto más, O, al conjunto $E(\mathbb{Q})$ de soluciones racionales, jugando el papel de *foco común* de todas las rectas verticales. Siguiendo la misma notación que en el apartado 2.1, podríamos llamar

$$\bar{E}(\mathbb{Q}) := E(\mathbb{Q}) \cup \{O\}$$

el conjunto de *puntos racionales proyectivos* de E, aunque lo cierto es que tradicionalmente uno comete un pequeño abuso de notación y simplemente denota por $E(\mathbb{Q})$ al conjunto de puntos racionales de E ampliado con O. A partir de ahora, supondremos que nuestra curva elíptica viene dada por una ecuación de Weierstrass como en (38) y que el punto O es un punto más de $E(\mathbb{Q})$, ¡quizás el único!

En el caso de la curva elíptica (30) definida por la ecuación de Weierstrass $y^2 = x^3 - x$, por ejemplo, este abuso de notación nos lleva a convenir que

$$E(\mathbb{Q}) = \{O, (-1, 0), (0, 0), (1, 0)\}. \tag{39}$$

3.1. La estructura algebraica de $E(\mathbb{Q})$. Es importante que nos entretengamos en aclarar qué se entiende por *estructura algebraica* de $E(\mathbb{Q})$. ¡Hasta el momento para nosotros $E(\mathbb{Q})$ no era más que un conjunto sin ningún tipo de organización interna!

La idea fundamental que permite dotar al conjunto $E(\mathbb{Q})$ de una estructura algebraica razonable es muy natural si nos remontamos a la sección anterior. Allí vimos que cuando tomamos un punto racional $P_1 = (x_1, y_1)$ cualquiera, la recta tangente a E en P_1 corta la curva –salvo casos excepcionales– en tres puntos *racionales*: el mismo P_1 contado dos veces y un tercer punto, que en ese momento llamamos (x_3, y_3). Que (x_3, y_3) es nuevamente una solución racional se razonó en el apartado 2.3 mediante un argumento básico pero crucial:

Si un polinomio $p(x)$ de grado 3 y coeficientes racionales tiene 2 raíces racionales, entonces la tercera raíz también debe ser racional.

En esa situación, $p(x)$ era el polinomio de intersección de la recta tangente con la curva, y dos de las raíces de $p(x)$ eran la coordenada x_1 contada con multiplicidad dos.

Ahora nos gustaría aplicar el mismo principio a una construcción levemente diferente a ésta, que nos dará más libertad de acción: si en lugar de fijar *un* punto racional $P_1 = (x_1, y_1)$ en $E(\mathbb{Q})$, tomamos *dos*, $P_1 = (x_1, y_1)$ y $P_2 = (x_2, y_2)$, entonces podemos trazar la recta R que los une (entendiendo que R es la recta tangente si insistimos en tomar $P_1 = P_2$) y calcular el punto $P_3 = (x_3, y_3)$ que aparece como tercer punto de intersección entre R y E.

El citado principio nos permite concluir, como hicimos en el apartado 2.3, que

$$P_1, P_2 \in E(\mathbb{Q}) \;\Rightarrow\; P_3 \in E(\mathbb{Q}).$$

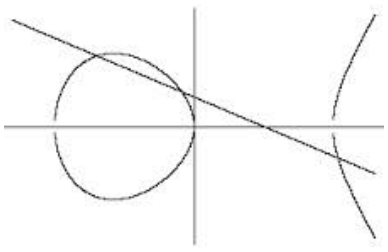


FIGURA 9. La recta que une P y Q corta además la curva elíptica en un tercer punto más.

Y aquí puede apreciar uno la gran ventaja que tiene el incluir el punto del infinito O como un punto racional más de $E(\mathbb{Q})$: la implicación anterior siempre

tiene sentido, incluso cuando la recta que une P_1 con P_2 es excepcional. Esto sucede cuando los dos puntos están alineados verticalmente y, como ya comenté, la recta que los une tiene a $P_3 = O$ como tercer punto de intersección.

La implicación anterior también es correcta cuando uno de los dos puntos base, pongamos P_1, es O: la recta que *une* O con P_2 es (¿lo adiviináis?) la recta vertical por $P_2 = (x_2, y_2)$, y por tanto el tercer punto de intersección P_3 es este caso el *otro* punto en que esa recta corta la curva, es decir, $P_3 = (x_2, -y_2)$.

Finalmente, hay un caso extremo que también tiene cabida: si $P_1 = P_2 = O$, decretamos que $P_3 = O$. La explicación geométrica de esta definición es la siguiente: en el plano proyectivo en el que se encuentra la curva E, la recta tangente a O pasa por el punto con multiplicidad 3; el punto O es pues un punto de inflexión y por tanto el tercer punto de intersección de la recta con la curva es O mismo.

En conclusión, tenemos delante nuestro una *operación* en el conjunto $E(\mathbb{Q})$. Por operación entendemos lo que es por ejemplo la operación *suma* o *multiplicación* en el conjunto $\mathbb{Z}$ de los números enteros. Dados dos elementos del conjunto, obtenemos un tercero operándolos. A cada operación que uno se inventa, le puede otorgar un símbolo para abreviarla: la suma la abreviamos $+$ y la multiplicación $\cdot$. Para la operación en $E(\mathbb{Q})$ que acabamos de construir usaremos el símbolo $'\oplus'$. Es decir, tenemos:

$$E(\mathbb{Q}) \times E(\mathbb{Q}) \longrightarrow E(\mathbb{Q}), \quad (P_1, P_2) \mapsto P_3 := P_1 \oplus P_2.$$

Si para un punto $P = (x, y) \in E(\mathbb{Q})$ definimos $-P := (x, -y)$, observad que $-P$ también es solución racional de la ecuación E y tenemos por tanto derecho a escribir $-P \in E(\mathbb{Q})$. Para que esta definición sea completa para todos los puntos de $E(\mathbb{Q})$, acordamos también que $-O = O$. Con esta notación, la discusión anterior se traduce en las siguientes propiedades:

$$P \oplus O = -P, \quad P \oplus (-P) = O, \quad O \oplus O = O. \tag{40}$$

Siempre que uno dispone de una operación en un conjunto, es natural preguntarse qué propiedades satisface. Por ejemplo, es obvio que la operación $\oplus$ es *conmutativa* o *abeliana*: da lo mismo $P_1 \oplus P_2$ que $P_2 \oplus P_1$, el orden no importa ya que P_3 sólo depende de la recta que pasa por P_1 y P_2.

En cambio, la operación $\oplus$ carece de *elemento neutro* en $E(\mathbb{Q})$: el candidato natural sería el punto O fijado, y para que actuara efectivamente como elemento neutro debería cumplirse que para todo $P \in E(\mathbb{Q})$:

$$\text{existiera algún } Q \in E(\mathbb{Q}) \text{ tal que } P \oplus Q = O, \text{ y } P \oplus O = P.$$

Hemos visto ya en (40) que la primera condición se cumple, tomando $Q = -P$. Pero (40) también nos muestra que la segunda condición *no* se satisface, aunque tiene fácil arreglo: podemos definir una nueva operación mediante la fórmula

$$P_1 + P_2 := -(P_1 \oplus P_2), \tag{41}$$

como se ilustra en las figuras 10 y 11.

Observad que la operación $+$ que hemos obtenido combinando las operaciones $\oplus$ y $-$ es de nuevo conmutativa. Y teniendo en cuenta que $-O = O$, deducimos

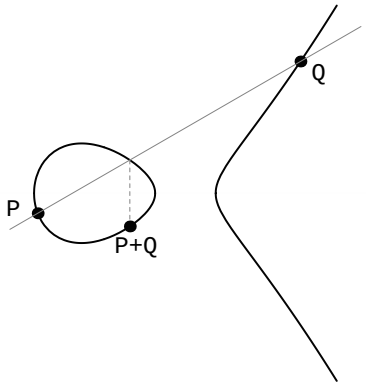


FIGURA 10. El resultado de operar P y Q con la operación $+$.

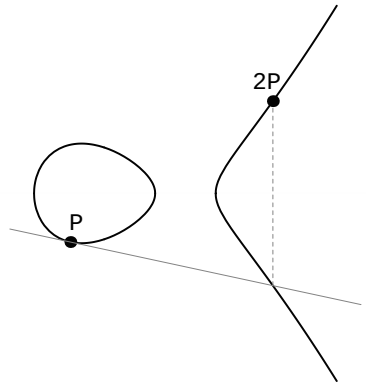


FIGURA 11. El resultado de la operación $P+P$, que simplemente denotamos $2P$.

de (40) que

$$P \oplus O = P, \quad P \oplus (-P) = O, \quad O \oplus O = O. \tag{42}$$

para todo punto $P \in E(\mathbb{Q})$, con lo que ahora sí O se comporta como elemento neutro para esta operación. Finalmente, aunque más arduo y pesado de demostrar, también es cierto que la operación $+$ definida en (42) es *asociativa*: para toda terna de puntos $P, Q, R \in E(\mathbb{Q})$ se cumple que $(P + Q) + R = P + (Q + R)$.

Cuando un conjunto está dotado de una operación conmutativa y asociativa y contiene un elemento neutro respecto la operación, se dice que es un *grupo conmutativo* o *grupo abeliano*. Los ejemplos más clásicos de grupos abelianos son cualquiera de los conjuntos de números $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$, con la operación $+$ usual y

el número 0 actuando como elemento neutro. A esta lista de grupos abelianos podemos incorporar ahora $E(\mathbb{Q})$, el *grupo abeliano de puntos racionales* de cualquier curva elíptica E.

Definición 3.2.
- *Dado un punto $P \in E(\mathbb{Q})$ y un número entero positivo $n > 0$, ponemos*
$$n \cdot P = P + \overset{n}{\ldots} + P$$
y
$$-n \cdot P = (-P) + \overset{n}{\ldots} + (-P).$$
Ponemos también $0 \cdot P = O$.
- *El orden de un punto $P \in E(\mathbb{Q})$ es el menor número $n > 0$ tal que $n \cdot P = O$, y lo denotamos $\mathrm{ord}(P) = n$. En particular, $\mathrm{ord}(O) = 1$. Si no existe ningún $n > 0$ tal que $n \cdot P = O$, entonces decimos que P tiene orden infinito y lo denotamos $\mathrm{ord}(P) = \infty$.*
- *El conjunto de puntos de torsión de $E(\mathbb{Q})$ es*
$$E(\mathbb{Q})_{tors} := \{\, P \in E(\mathbb{Q}),\ \mathrm{ord}(P) \ \text{ es finito} \,\}.$$

En el caso de la curva elíptica $E : y^2 = x^3 - x$, podéis comprobar que los tres puntos $P = (-1, 0)$, $Q = (0, 0)$ y $R = (1, 0)$ cumplen

$$2 \cdot P = O,\ \ 2 \cdot Q = O,\ \ 2 \cdot R = O,\ \ P + Q = R,\ \ P + R = Q,\ \ Q + R = P,$$

lo cual determina completamente la estructura de grupo en $E(\mathbb{Q})$. En particular, P, Q y $R \in E(\mathbb{Q})$ son puntos de torsión, concretamente de orden 2.

Esto nos permite darnos cuenta que $E(\mathbb{Q})$ no es ningún grupo extraño, sino más bien un grupo bastante simple y familiar. Concretamente, si $\mathbb{Z}/2\mathbb{Z} = \{\bar{0}, \bar{1}\}$ denota el grupo de dos elementos con la operación $+$ definida por las reglas

$$\bar{0} + \bar{0} = \bar{0}, \quad \bar{0} + \bar{1} = \bar{1}, \quad \bar{1} + \bar{0} = \bar{1}, \quad \bar{1} + \bar{1} = \bar{0}, \tag{43}$$

la aplicación

$$\begin{array}{ccc} E(\mathbb{Q}) = \{O, (-1,0), (0,0), (1,0)\} & \longrightarrow & \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} = \{(\bar{0},\bar{0}), (\bar{0},\bar{1}), (\bar{1},\bar{0}), (\bar{1},\bar{1})\} \\ O & \mapsto & (\bar{0},\bar{0}) \\ P & \mapsto & (\bar{0},\bar{1}) \\ Q & \mapsto & (\bar{1},\bar{0}) \\ R & \mapsto & (\bar{1},\bar{1}) \end{array}$$

proporciona un *isomorfismo de grupos* $E(\mathbb{Q}) \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. En otras palabras, como grupos abstractos, ¡$E(\mathbb{Q})$ y $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ son el mismo!

Por otro lado, al final de la sección 1.6 vimos que la curva elíptica (36) de ecuación de Weierstrass $y^2 = x^3 + 17$ tiene muchos más puntos racionales, como $P_1 = (-1, 4)$, que tiene coordenadas enteras y otros, que allí denotamos P_2, P_3 y P_4, de coordenadas racionales no-enteras que no volvemos a reproducir aquí porque el número de dígitos de sus numeradores y denominadores es bastante grande. Si uno lleva a cabo los cálculos (siguiendo los mismos pasos que los realizados en la discusión que precede al teorema 2.2), puede comprobarse fácilmente que

$$P_2 = -2P_1, \quad P_3 = 3P_1 \quad \text{y} \quad P_4 = 4P_1;$$

esto muestra cómo se calculó el punto P_3 de la sección 1.6, e ilustra la siguiente propiedad, inmediata a partir de las definiciones:

Proposición 3.3. *Sea $P_1 = (x_1, y_1) \in E(\mathbb{Q})$ un punto racional en una curva elíptica. El punto $P_3 = (x_3, y_3)$ que se obtiene al aplicar el algoritmo de la tangente a P_1 es*

$$P_3 = -2P_1. \tag{44}$$

La fórmula (44) también explica el porqué del teorema 2.2 y la proposición 2.3: la sucesión de puntos racionales que se obtiene al iterar indefinidamente el algoritmo de la recta tangente es

$$P_1, -2P_1, 4P_1, -8P_1, 16P_1, -32P_1, ...$$

que, si $\text{ord}(P) = \infty$, proporciona una cantidad infinita de puntos diferentes, y a su vez pasa por alto una cantidad también infinita de puntos de $E(\mathbb{Q})$: los $2P_1$, $\pm 3P_1$, $-4P_1$, $\pm 5P_1$, etc.

En los años veinte, L. J. Mordell demostró el siguiente resultado fundamental sobre $E(\mathbb{Q})$, que, aunque no podemos demostrar aquí, su enunciado permite hacernos una idea más clara de la estructura interna de este grupo. En la versión que presento aquí añado un dato adicional, nada trivial, sobre el tamaño de $E(\mathbb{Q})_{tors}$ que descubrió el matemático B. Mazur más de cincuenta años después:

Teorema 3.4. (a) *$E(\mathbb{Q})_{tors}$ es un subgrupo* finito *de $E(\mathbb{Q})$ de tamaño menor o igual que* 16.

(b) *Existen un número $r \geq 0$ y puntos $Q_1, ..., Q_r \in E(\mathbb{Q})$ de orden infinito tales que todo punto $Q \in E(\mathbb{Q})$ puede expresarse de manera única como*

$$Q = n_1 Q_1 + ... + n_r Q_r + T$$

donde $n_1, ..., n_r \in \mathbb{Z}$ son números enteros y $T \in E(\mathbb{Q})_{tors}$ es un punto de torsión.

En dos palabras, el teorema de Mordell afirma que el grupo abeliano $E(\mathbb{Q})$ está *finitamente generado.*

Definición 3.5. *El número r que aparece en el enunciado del teorema de Mordell se llama el* rango *de $E(\mathbb{Q})$.*

En el caso de la curva elíptica de ecuación $y^2 = x^3 - x$, hemos comprobado ya que $E(\mathbb{Q}) = E(\mathbb{Q})_{tors}$ y por tanto su rango es $r = 0$.

La curva elíptica (36), de ecuación $E. y^2 = x^3 + 17$, es más jugosa. Fijaos en que en la sección 1.6 mostramos *diecinueve* puntos diferentes de $E(\mathbb{Q})$: los dieciséis con coordenadas enteras y los tres puntos adicionales P_2, P_3 y P_4. A la vista del apartado (a) anterior, esto significa que no todos ellos son de torsión, con lo que debe haber alguno de orden infinito y por tanto el rango de E es como mínimo 1.

Con técnicas más avanzadas, basadas en los mismos argumentos necesarios para demostrar el teorema de Mordell, se puede comprobar que en realidad ¡*ninguno* de los puntos de $E(\mathbb{Q})$ –salvo O, claro está– es de torsión! Más precisamente, el rango de E es $r = 2$ y $E(\mathbb{Q})$ está generado por los puntos $(-1, 4)$ y $(-2, 3)$.

El teorema de Mordell nos dice que la estructura de grupo de $E(\mathbb{Q})$ –lo que al principio de esta sección llamábamos la *estructura algebraica* de $E(\mathbb{Q})$– queda completamente determinada si conocemos el grupo finito $E(\mathbb{Q})_{tors}$ y el rango r. Calcular $E(\mathbb{Q})_{tors}$ no conlleva habitualmente ninguna dificultad: la aportación de Mazur reduce el número de posibles estructuras de grupo en $E(\mathbb{Q})_{tors}$ a una lista finita, y decidir cuál es acostumbra a ser una tarea fácil.

Por el contrario, el rango r de $E(\mathbb{Q})$ es un invariante mucho más misterioso, del que da cuenta la conjetura de Birch y Swinnerton-Dyer, como veremos en las páginas siguientes.

3.2. La función $L_E(s)$. El capítulo anterior de este volumen está dedicado a la hipótesis de Riemann sobre la función analítica

$$\zeta : \{s \in \mathbb{C},\, \mathrm{Re}(s) > 1\} \longrightarrow \mathbb{C}, \quad \zeta(s) := \sum_{n \geq 1} \frac{1}{n^s} = \prod_p \frac{1}{(1 - p^{-s})}, \tag{45}$$

conocida como la función zeta de Riemann. En la suma, n recorre todos lo números naturales, mientras que en el producto p recorre todos los números primos $2, 3, 5, 7, 11...$

Aunque la suma y el producto sólo convergen cuando la parte real $\mathrm{Re}(s)$ de la variable compleja s es mayor que 1, se vio que la función puede extenderse a una función meromorfa en todo el plano complejo, con un solo polo en el punto $s = 1$. La función resultante, $\zeta : \mathbb{C} \setminus \{1\} \longrightarrow \mathbb{C}$, es el paradigma clásico de función zeta o función L, como muchos otros la llaman, que se suele asociar a un objeto aritmético.

Uno encuentra dispersas en la literatura funciones $L_M(s)$ asociadas a todo tipo de objetos M, todas ellas definidas bajo el mismo patrón, aunque no siempre tratadas como parte de una teoría general que las engloba a todas. Concretamente, siempre se tiene

$$L_M(s) := \prod_{p \nmid \Delta(M)} \frac{1}{L_p(p^{-s})} \tag{46}$$

donde $\Delta(M)$ es un número entero no nulo llamado el *discriminante* de M, p recorre el conjunto de todos los números primos que no dividen a $\Delta(M)$ y, para cada uno de ellos, $L_p(T)$ es un polinomio con coeficientes enteros y de grado $d \geq 1$ fijo, que llamamos el *grado* de M. En la fórmula (46), $L_p(p^{-s})$ no es otra cosa que el valor que toma el polinomio $L_p(T)$ cuando hacemos la substitución $T = p^{-s}$.

La receta usada para definir el polinomio $L_p(T)$ es también la misma en todos los casos: $L_p(T)$ es el polinomio característico de un cierto endomorfismo Fr_p, comúnmente llamado el endomorfismo de Frobenius por el matemático alemán que lo introdujo, actuando en un cierto espacio vectorial de dimensión d asociado a M que recibe varios nombres según el contexto, aunque en general es conocido como el espacio de cohomología étale de M.

En el caso de la función zeta de Riemann, el discriminante es $\Delta(\mathbb{Z}) = 1$, el grado es $d = 1$ y para todo primo p, el polinomio $L_p(T)$ es el polinomio característico

del endomorfismo *identidad* actuando en $\mathbb{Z}$, que no es otro que $L_p(T) = 1 - T$. Observad que con estos ingredientes efectivamente se obtiene (45).

Aunque no definiremos aquí qué son la mayoría de los objetos que mencionamos, listo a continuación otras funciones L que están recibiendo mucha atención en la actualidad: podéis consultar las referencias sugeridas para más detalles, y sirva sino esta lista como catálogo de objetos aritméticos sumamente interesantes que están relacionados entre sí por una densa red de analogías y conjeturas:

- La función zeta de Riemann (45), que se considera la función L asociada al grupo $\mathbb{Z}$ y por tanto podríamos denotar $\zeta(s) = L_{\mathbb{Z}}(s)$.
- La función zeta de Dedekind $\zeta_K(s)$ de un cuerpo de números, que se define de manera muy similar a (45) substituyendo números primos por ideales primos, y puede interpretarse como la función $L_{\mathcal{O}_K}(s)$ del anillo de enteros de K; véase [16].
- La función zeta o L de Hasse-Weil $L_E(s)$ asociada a una curva elíptica, que veremos a continuación, y sus generalizaciones en dimensión superior; véase [3], [17] o [24].
- La función $L_\varrho(s)$ asociada a una representación de Galois artiniana, o a un sistema compatible de representaciones de Galois ℓ-adicas; véase [11], [19].
- La función $L_f(s)$ asociada a una forma modular y, más generalmente, la función $L_\pi(s)$ asociada a una representación automorfa; véase [20], [11].

El producto (46) que define todas esta funciones converge y define una función holomorfa en un subconjunto abierto del plano complejo de la forma $\{s \in \mathbb{C}, \mathrm{Re}(s) > s_0\}$ para algún número real s_0. Para todas ellas se sabe o se espera que puedan extenderse a una función meromorfa en todo el plano complejo.

Finalmente, para varias de estas funciones L existe una versión de la conjetura de Birch y Swinnerton-Dyer.

En los dos primeros casos, que son los más clásicos, en realidad no se trata de una conjetura sino de un teorema demostrado, conocido como la *fórmula analítica del número de clases* y que podéis consultar en [16]. Esta fórmula fue sin duda una de las fuentes de inspiración para los matemáticos B. Birch y P. Swinnerton-Dyer al plantear su conjetura para curvas elípticas. Finalmente, versiones mucho más generales de esta conjetura se deben a matemáticos como H. Stark, A. Beilinson, S. Bloch, K. Kato y otros.

En general, a qué objetos aritméticos se puede, se sabe o se espera asignar una función L es una pregunta que ha sido durante muchos años y sigue siendo fuente de debate, y hasta se ha acuñado un término para denominarlos: *motivos*. Así, $\mathbb{Z}$, el anillo de enteros de un cuerpo de números, las curvas elípticas, las representaciones de Galois y las formas modulares dan lugar todos ellos a un objeto en la categoría común de los motivos. Podéis consultar [1], [12] y [15] para una introducción a esta teoría.

Habiendo introducido ya el contexto natural en el que debe situarse, estamos en condiciones de definir la función L de una curva elíptica

$$E:\ y^2 = x^3 + Ax + B, \tag{47}$$

donde A y B son coeficientes enteros tales que $\Delta(E) = -16(4A^3 + 27B^2) \neq 0$.

Para definirla, seguiremos el patrón general dado en (46). El discriminante de E no es otro que $\Delta(E)$, y el grado será $d = 2$. Para definir el polinomio $L_p(T)$ de grado 2 asociado a un primo $p \nmid \Delta(E)$ fijado, necesitamos introducir el conjunto de los números enteros módulo p.

Dado un número entero $a \in \mathbb{Z}$, definimos $\bar{a}$ como el resto de la división entera de a entre p, que llamamos la *clase de a módulo p*. Es por tanto un número entre 0 y $p-1$. Como hicimos ya para $p = 2$ en (43), podemos considerar el conjunto

$$\mathbb{Z}/p\mathbb{Z} = \{\bar{0}, \bar{1}, ..., \overline{p-1}\}$$

de los *números enteros módulo p*, que podemos dotar con dos operaciones $+$ y $\cdot$ del modo siguiente: dados dos elementos $\bar{a}$, $\bar{b}$ en $\mathbb{Z}/p\mathbb{Z}$, ponemos

$$\bar{a} + \bar{b} := \bar{r},$$

donde r es el resto de la división entera de $(a+b)$ entre p, y

$$\bar{a} \cdot \bar{b} := \bar{r},$$

donde r es el resto de la división entera de $(a \cdot b)$ entre p.

Si esta es la primera vez que veis estas operaciones, a lo mejor os parecerán extrañas, pero es algo sencillo y natural. Por ejemplo, si $p = 7$, tenemos

$$\mathbb{Z}/7\mathbb{Z} = \{\bar{0}, \bar{1}, ..., \bar{6}\}$$

y $\bar{5} + \bar{4} = \bar{2}$, porque el resto de la división de $5 + 4 = 9$ entre 7 es igual a 2. Similarmente, $\bar{5} \cdot \bar{4} = \bar{6}$ porque el resto de la división de $5 \cdot 4 = 20$ entre 7 es 6.

De esta manera, en el conjunto $\mathbb{Z}/p\mathbb{Z}$ podemos sumar y multiplicar elementos, como lo hacemos también en los conjuntos $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ y $\mathbb{C}$.

Definición 3.6. *Sea E una curva elíptica como en* (47) *y p un número primo que no divida a $\Delta(E)$. Definimos el conjunto de* puntos enteros módulo *p de E como*

$$E(\mathbb{Z}/p\mathbb{Z}) = \{O\} \cup \{(\bar{x}, \bar{y}) \in \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \text{ tales que } \bar{y}^2 = \bar{x}^3 + \bar{A} \cdot \bar{x} + \bar{B} \text{ en } \mathbb{Z}/p\mathbb{Z}\}.$$

Al contrario de $E(\mathbb{Q})$, disponemos de un algoritmo sencillo para calcular el conjunto $E(\mathbb{Z}/p\mathbb{Z})$ explícitamente en un número finito de pasos: además de O, que siempre lo contamos como punto de $E(\mathbb{Z}/p\mathbb{Z})$, sólo hay que decidir si cada uno de los p^2 puntos $(\bar{x}, \bar{y})$ en $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ satisface la ecuación (47) o no.

Aunque no entraremos en detalles aquí, éste no es el método más óptimo para determinar $E(\mathbb{Z}/p\mathbb{Z})$ y existen maneras más inteligentes de hacerlo: de entre los p^2+1 puntos diferentes que en principio podría haber en $E(\mathbb{Z}/p\mathbb{Z})$, un resultado de H. Hasse asegura que en $E(\mathbb{Z}/p\mathbb{Z})$ hay muchos menos. Concretamente, el tamaño de $E(\mathbb{Z}/p\mathbb{Z})$ es

$$\sharp E(\mathbb{Z}/p\mathbb{Z}) = p + 1 - a_p \tag{48}$$

donde $a_p \in \mathbb{Z}$ es un número entero que está convenientemente acotado:

$$-2\sqrt{p} \leq a_p \leq 2\sqrt{p}. \tag{49}$$

El número a_p a veces se llama el término de error de Hasse, y otros lo conocen como la *traza del endomorfismo de Frobenius en p* porque a_p puede interpretarse también como eso: la traza de un cierto endomorfismo Fr_p –al que no hemos referido antes también– actuando en un cierto espacio vectorial de dimensión $d = 2$, en este caso tradicionalmente conocido como el *módulo de Tate* de E.

Para más información, podéis consultar [21]. En cualquier caso, se tiene $L_p(T) = 1 - a_pT + pT^2$ y por tanto:

Definición 3.7. *La función L de E es*

$$L_E(s) = \prod_{p \nmid \Delta(E)} \frac{1}{1 - a_p \cdot p^{-s} + p \cdot p^{-2s}}. \tag{50}$$

Las desigualdades (49) de Hasse permiten mostrar, de manera similar al resultado análogo para la función zeta de Riemann, que este producto converge cuando la parte real de la variable compleja s es mayor que $\frac{3}{2}$, y da pie a una función holomorfa de variable compleja $L_E : \{s \in \mathbb{C},\, \mathrm{Re}(s) > \frac{3}{2}\} \longrightarrow \mathbb{C}$.

Que también puede extenderse a todo el plano complejo es cierto, pero sólo se sabe demostrar como consecuencia de unos de los teoremas más importantes de la historia de la teoría de números: el teorema de modularidad del matemático A. Wiles, publicado en 1995 en [25]. El teorema de Wiles afirma que

$$L_E(s) = L_f(s) \tag{51}$$

para una cierta *forma modular* f, concepto que no desarrollaremos aquí y para el cual referimos a [5].

Comoquiera que sea, la función $L_f(s)$ asociada a una forma modular es una función mucho más tratable que la de una curva elíptica: entre otras propiedades, se sabe demostrar por métodos elementales que $L_f(s)$ puede extenderse a una función *holomorfa* en todo $\mathbb{C}$, sin polo alguno. Al combinar este resultado con (51), se deduce lo propio de $L_E(s)$, obteniendo así una función holomorfa

$$L_E : \mathbb{C} \longrightarrow \mathbb{C}.$$

3.3. La conjetura de Birch y Swinnerton-Dyer. ¿Qué relación guarda la función $L_E(s)$ de una curva elíptica E con el rango r de su grupo $E(\mathbb{Q})$ de puntos racionales?

En los años sesenta, los matemáticos B. Birch y P. Swinnerton-Dyer se dejaron llevar por el siguiente razonamiento (o heurística, como a veces se llama), de dudosa validez pero muy sugerente: aunque *el producto* (50) que define la función $L_E(s)$ *diverge* en el punto $s = 1$, la expresión

$$\text{"}L_E(1)\text{"} := \prod_{p \nmid \Delta(E)} \frac{1}{1 - a_p \cdot p^{-1} + p \cdot p^{-2}} = \prod_{p \nmid \Delta(E)} \frac{p}{p - a_p + 1} = \prod_{p \nmid \Delta(E)} \frac{p}{\sharp E(\mathbb{Z}/p\mathbb{Z})}$$

que se obtiene al substituir *formalmente* $s = 1$ en (50) debería guardar alguna relación con el valor *real* que toma la función (extendida) $L_E(s)$ en $s = 1$.

Notad que en esos años tal interpretación era incluso más ilegítima que ahora: la continuidad analítica de $L_E(s)$ a todo el plano complejo se ignoraba por completo, ya que por ese entonces el teorema de modularidad (51) de Wiles era sólo una controvertida conjetura, propuesta por los matemáticos G. Shimura, Y. Taniyama y A. Weil.

Por otro lado, si se invierte la expresión formal que encontramos para "$L_E(1)$", esto nos conduce a estudiar el comportamiento asintótico de la función

$$\mathrm{BSD}_E : \mathbb{N} \longrightarrow \mathbb{R}, \quad \mathrm{BSD}_E(x) = \prod_{p<x} \frac{\sharp E(\mathbb{Z}/p\mathbb{Z})}{p}$$

cuando x tiende a infinito. De nuevo parece razonable sospechar que la siguiente cadena de implicaciones debería guardar algo de cierto:

Cuanto mayor es el rango r de $E(\mathbb{Q})$, más puntos racionales hay en $E(\mathbb{Q})$ y por tanto más puntos deberíamos encontrar en $E(\mathbb{Z}/p\mathbb{Z})$ al variar p. En consecuencia, la función BSD_E *debería crecer más rápidamente cuando x tiende a infinito.*

Dicho y hecho, B. Birch y P. Swinnerton-Dyer se pusieron manos a la obra: tomaron varias curvas elípticas E cuyo rango r conocían de antemano, como la curva (30) de rango $r = 0$ o la curva (36) de rango 2 y representaron gráficamente la función BSD_E en el intervalo mayor posible de la variable x.

Las gráficas que se obtienen hoy en día para $x \in [1, 50000]$ son las siguientes:

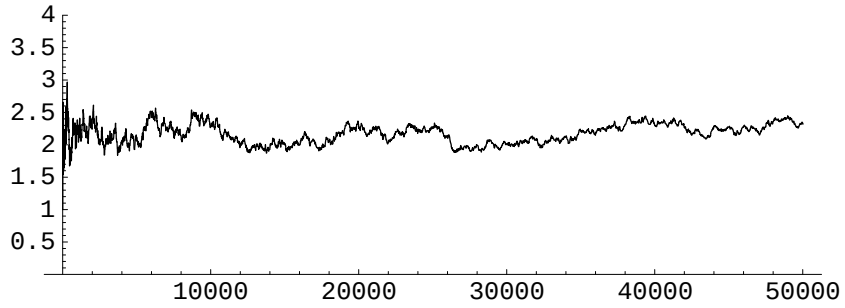


FIGURA 12. La función BSD_E para la curva elíptica $E : y^2 = x^3 - x$ de rango $r = 0$. Su comportamiento asintótico es comparable al de una función constante $f(x) = A$, donde $A \sim 2$.

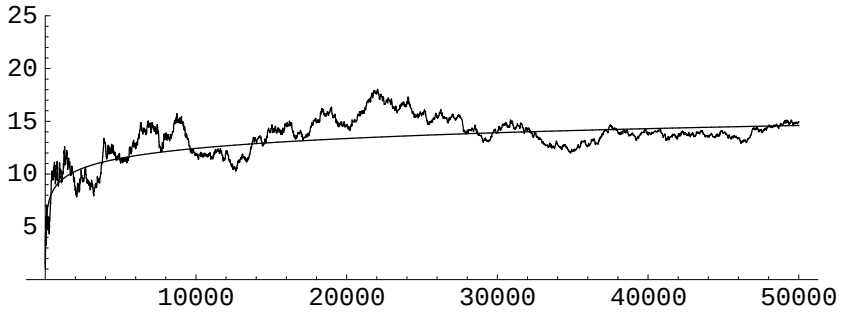


FIGURA 13. La función BSD_E para la curva elíptica $E : y^2 = x^3 - 5x$ de rango $r = 1$. Su comportamiento asintótico es comparable al de la función $f(x) = A \cdot \log(x)$ para un valor adecuado de la constante A.

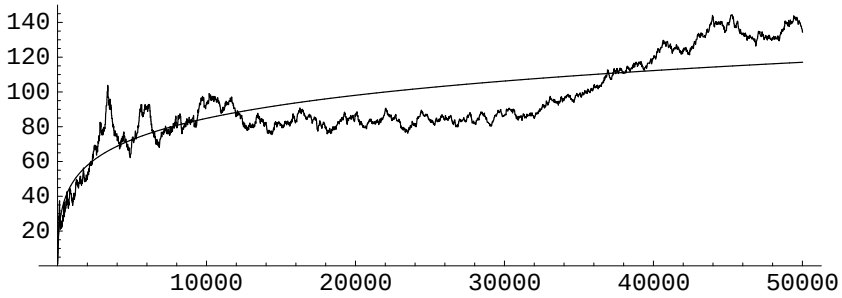


FIGURA 14. La función BSD_E para la curva elíptica $E : y^2 = x^3 - 17x$ de rango $r = 2$. Su comportamiento asintótico es comparable al de la función $f(x) = A \cdot \log^2(x)$ para un valor adecuado de la constante A.

Estos cálculos llevaron a los dos matemáticos a formular la siguiente conjetura:

Conjetura 3.8 (Birch y Swinnerton-Dyer)**.** *Sea E una curva elíptica y sea r su rango. El comportamiento asintótico de la función* BSD_E *cuando x tiende a infinito es igual al de la función $f(x) = A \cdot \log^r(x)$ para un valor adecuado de la constante A.*

Cuando uno revisa con atención la relación precisa entre la función BSD_E y el valor *real* de la función $L_E(s)$ en $s = 1$ (y podéis consultar los detalles de dicho cálculo en [2] y [8]), obtiene la siguiente formulación equivalente de la conjetura anterior, que es la que aparece oficialmente en el sitio web del Clay Mathematics Institute:

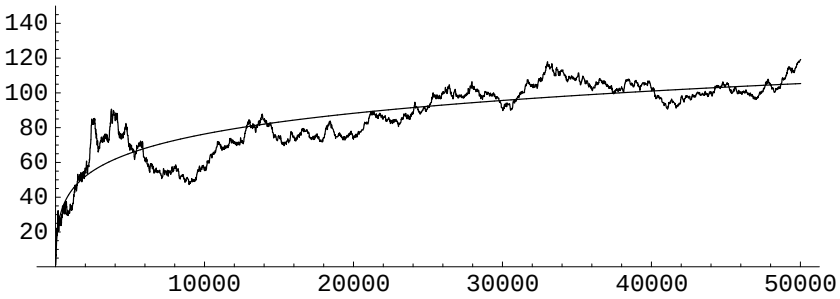


FIGURA 15. La función BSD_E para la curva elíptica $E : y^2 = x^3 - 56x$ de rango $r = 3$. Su comportamiento asintótico es comparable al de la función $f(x) = A \cdot \log^3(x)$ para un valor adecuado de la constante A.

Conjetura 3.9 (Birch y Swinnerton-Dyer)**.** *Sea E una curva elíptica y sea $r = r(E)$ su rango. Entonces el desarrollo de Taylor de la función $L_E(s)$ alrededor del punto $s = 1$ es de la forma*

$$c(s-1)^r + \text{ términos de orden superior}$$

para alguna constante $c \neq 0$.

En otras palabras, si denotamos por $r_{an}(E)$ el orden de anulación de la función $L_E(s)$ en $s = 1$, la conjetura de Birch y Swinnerton-Dyer predice que

$$r(E) \stackrel{?}{=} r_{an}(E). \tag{52}$$

El resultado más importante en relación a la conjetura es el siguiente teorema, que surge como combinación de los trabajos de V. Kolyvagin, B. H. Gross, D. Zagier y S. Zhang que se pueden consultar en [9], [10], [14] y [26].

Teorema 3.10. *Sea E una curva elíptica.*

- *Si $r_{an}(E) = 0$, entonces $r(E) = 0$.*
- *Si $r_{an}(E) = 1$, entonces $r(E) = 1$.*

La demostración de este resultado está completamente fuera de las posibilidades de este texto. Aparte de los artículos originales de los autores del teorema, se pueden consultar también otros tratados como [3] ó [4], que pueden considerarse como una continuación natural de los conceptos introducidos en este capítulo, aunque dirigidos ya a un lector con un dominio de los resultados básicos de la teoría de curvas elípticas y curvas modulares como los desarrollados, por ejemplo, en [5], [21] y [22].

4. Conclusión

En el momento de escribir estas líneas, la conjetura de Birch y Swinnerton-Dyer permanece completamente abierta. Parece obvio que cualquier demostración que

se pueda dar de la conjetura requiere una idea revolucionaria, diferente de todos los puntos de vista que se han tomado hasta ahora para aproximarse a ella.

Pero el interrogante que plantea no es de esos de *todo o nada*: uno puede ser menos ambicioso y considerar casos particulares de la conjetura que ofrecen ya retos importantes, nada triviales pero quizás más asequibles. Así, la conjetura de Birch y Swinnerton-Dyer es fuente de numerosos problemas matemáticos de gran interés, cuya resolución son pequeños pasos hacia la demostración de la conjetura completa.

De esta manera, aunque uno puede sentirse aún muy lejos de conocer la ciudad entera, hay centenares de rincones por explorar que merecen por sí mismos ser investigados.

Ilustraciones: Maria Vidal

Referencias

[1] J. Coates, On p-adic L-functions, *Séminaire Bourbaki* **701** (1988-1989), 33-59.

[2] K. Conrad, Partial Euler products on the critical line, *Canad. J. Math.* **57** (2005), 267–297.
[3] H. Darmon, Rational points on modular elliptic curves, *CBMS Regional Conference Series in Mathematics* **101**, American Mathematical Society, 2004.
[4] H. Darmon, V. Rotger, Algebraic cycles and Stark-Heegner points, notas de un curso impartido en la *Arizona Winter School on Stark-Heegner points*, 2011. Disponibles en las páginas web de los autores.
[5] F. Diamond, J. Shurman, A first course in modular forms, *Grad. texts Math.* **228**, Springer Verlag, 2005.
[6] G. Faltings, Endlichkeitssätze für abelsche Varietäten über Zahlkörpern, *Invent.Math.* **73** (1983), 349-366.
[7] W. Fulton, Curvas algebraicas, *Editorial Reverté*, traducción de J. Pla, 2005.
[8] D. Goldfeld, Sur les produits partiels eulériens attachés aux courbes elliptiques, *C. R. Acad. Sci. Paris Sr. I Math.* **294** (1982), 471-474.
[9] B. H. Gross, D. B. Zagier, Heegner points and derivatives of L-series, *Invent. Math.*, **84:2** (1986), 225–320.
[10] B. H. Gross, Kolyvagin's work on modular elliptic curves in *L-functions and arithmetic* (Durham, 1989), 235–256, *London Math. Soc. Lecture Note Ser.* **153**, Cambridge Univ. Press, Cambridge, 1991.
[11] H. Iwaniec, E. Kowalski, Analytic Number Theory, *Colloquium Publications* **53**, Amer. Math. Soc., 2004.
[12] M. Kim, An introduction to motives I: classical motives and motivic L-functions, notas de un ciclo de conferencias organizadas en la *IHES summer school on motives*, 2006. Disponibles en la página web del autor.
[13] N. Koblitz, A Course in Number Theory and Cryptography, *Grad. texts Math.* **114**, Springer, 1987.
[14] V.A. Kolyvagin, Finiteness of $E(\mathbf{Q})$ and $\underline{\text{Ш}}(E,\mathbf{Q})$ for a subclass of Weil curves, *Izv. Akad. Nauk SSSR Ser. Mat.* **52** (1988), 670–671; traducción al inglés en *Math. USSR-Izv.* **32** (1989), 523–541.
[15] B. Mazur, What is a Motive?, *Notices Amer. Math. Soc.* **51:10** (2004), 1214–1216.
[16] J. Neukirch, Algebraic number theory, *Grundlehren der mathematischen Wissenschaften* **322**, 1999.
[17] J. S. Milne, Elliptic curves, *Booksurge Publishing*, 2006.
[18] P. Samuel, Projective Geometry, *Undergrad. texts Math.*, Springer, 1988.
[19] J.-P. Serre, Abelian ℓ-adic representations and elliptic curves, *W.A. Benjamin Inc.*, 1968.
[20] G. Shimura, Introduction to the Arithmetic Theory of Automorphic functions forms, *Princeton Univ. Press*, Princeton, N.J., 1971.
[21] J. H. Silverman, The arithmetic of elliptic curves, *Grad. Texts Math.* **106**, Springer Verlag.
[22] J. H. Silverman, Advanced topics in the arithmetic of elliptic curves, *Grad. Texts Math.*, Springer Verlag, 1994.
[23] J.H. Silverman, J. Tate, Rational Points on Elliptic Curves, *Undergrad. texts Math.*, Springer, 1992.
[24] J. Tate, Algebraic cycles and poles of zeta functions, in *Arithmetical Algebraic Geometry*, Harper and Row, 93-110, 1963.
[25] A. Wiles, Modular elliptic curves and Fermat's last theorem, *Annals Math.* (2) **141** (1995), 443–551.
[26] S. Zhang, Heights of Heegner points on Shimura curves. *Annals Math.* (2) **153** (2001), no. 1, 27–147.

Departament de Matemàtica Aplicada II, Universitat Politècnica de Catalunya, C. Jordi Girona 1-3, 08034 Barcelona

E-mail address: victor.rotger@upc.es

URL: http://www-ma2.upc.edu/vrotger